PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH

DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

ANNUAL REPORTS

Vol. I

FY 1967-71

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY
Report of Program Activities
July 1, 1966, through June 30, 1967

ANNUAL REPORT

TABLE OF CONTENTS

Page

July 1, 1966, through June 30, 1967

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Division Activities

1. DCRT-1
Serial Number

2. DIVISION OF COMPUTER RESEARCH & TECHNOLOGY

3. Dr. A. W. Pratt
Director, DCRT

I. SUMMARY

During this, the second full year of operation for the Division of Computer Research and Technology, significant progress has been made in 1) selecting and appointing men exceedingly well-qualified to conduct the research and operational programs of the Division, 2) converting the central computing facility's equipment to a third generation computer system and, 3) designing and developing new and imaginative applications of computers in support of biomedical research.

II. DIVISION PROGRAMS

A. Objectives

The objective of the Division is to plan and conduct an integrated computer research and service program in support of the NIH mission by 1) providing professional programming, computational and automatic data processing capability to meet NIH program needs, 2) conducting theoretical and applied research in mathematical statistics, mathematics and other physical and life sciences, 3) providing resources for research, development and consultation for the design and implementation of project-supporting computer systems, 4) providing scientific and administrative direction in the formulation of NIH-wide policies, standards, methods, and procedures on computation and automatic data processing activities, and, 5) participating in the analysis of requirements, design, and development of automatic data processing systems to make effective use of advanced techniques and equipment.

B. Summary of Program Progress and Accomplishments

1. Conversion to New Central Computer System

For almost two years now, the Division has been in the process of converting from Honeywell to IBM equipment. The conversion process has included both the installation and checking out of the equipment and its accompanying software as well as the reprogramming of Honeywell users programs. We have experienced exasperating delays in the delivery of equipment as well as the delivery of operational software. Likewise, progress in the conversion of Honeywell programs has been equally slow, far exceeding all estimates of the time required

for completion. However, the realization of complete conversion is within sight. Although the new equipment has not yet attained the desired speed and flexibility, it has greatly improved during this fiscal year. Hardware unreliability is still a source of concern and major steps are being taken by the vendor to correct this equipment deficiency.

The reprogramming effort on the part of IBM contractors is also proceeding satisfactorily with completion targeted for early Fall. The experiences of the past two years have had their effect on the DCRT program as well as on the configuration plans. The original goal of converting all work to an IBM System 360/65 has necessarily been modified to a more realistic configuration for the immediate future.

2. Central Computer Configuration

Because of the delays and uncertainties in the delivery and reliability of IBM equipment and software, the Division has modified its original plan to convert all work to a single 360 model 65 and instead is establishing a multi-machine configuration. With conversion completed by this Fall, the Division will operate a S-360/40 and two S-360/50's. The S-360/40 will be devoted to processing the DRG extramural awards. The remaining two S-360/50's will process the remaining NIH workload. The multi-machine configuration recognizes that the large NIH workload comprises widely different forms of computation and will make it possible to separate, for processing purposes, the mathematical computation from the data processing characteristics of collaborative studies and administrative applications. During the coming year the Division plans to add additional equipment as necessary to meet the increasing workload and provide teleprocessing capability for the widely ranging needs of scientists and administrators at NIH.

3. Reorganization and Staffing

The Division is undergoing a reorganization essential to the improved conduct of business. In January, 1) the Laboratory of Applied Studies was established for collaborative and in-house research in mathematics and mathematical statistics as applied to clinical research programs, 2) the Computer Systems Laboratory was established to aid and assist in the development of special-purpose biomedical computer applications in the Institutes and Divisions, and 3) the Physical Sciences Laboratory was established to conduct theoretical research in mathematics, physics and chemistry as part of developing the scientific basis for the application of computer technology to biomedicine.

The Computation and Data Processing Branch continues in its role of providing ADP service to the NIH. However, the Division is considering removing the EDP systems programing responsibility from the Branch and establishing it in a separate organization entity.

The Division has indeed been fortunate in recruiting excellent leadership to conduct its assigned research and development programs. With

the addition of a senior researcher this summer, around which a laboratory for research in heuristics and artificial intelligence will be formed, the Division will have completed its recruitment of a competent, top level research and development staff.

4. DCRT Library

In cooperation with the NIH Library, the Division established a library of books, periodicals, and journals pertaining to mathematics and the computer sciences. Although the library is only starting its collection, it has been well received by DCRT employees as well as numerous personnel throughout NIH.

5. Support Projects

The Division has continued to support Clinical Center automation during the year. Primary support has been directed toward Clinical Pathology automation (over one-half million dollars). The Division has purchased a small programmed console computer for the Radiation Branch, NCI. This computer will initially be used to compute isodose curves for radiation therapy planning. Later expansion of the system and linking it to the central facility will increase its utility.

6. Research Projects

The Division is young and the past year has been spent in defining, planning and organizing its research program. Completed projects are few; a substantial number of projects are underway and significant progress is being made as indicated in the summary reports of the individual laboratories.

C. Problems

1. Recruitment

Recruiting qualified programmers and computer operators continues as a perennial problem. Although the Civil Service Commission has made progress in refining the employment procedures, delays still exist which often result in the applicant seeking employment elsewhere.

2. Space

Although recruitment has been slow, it has considerably out-paced the acquisition of new space for the Division. Net employment will increase by at least fifty people during FY68 with no provision to house them. In addition, the Computer Systems Laboratory is in need of approximately 500 square feet of space in which to set up a laboratory essential in supporting the programs of the Institutes and Divisions.

July 1, 1966, through June 30, 1967

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT-2
Serial Number

2. COMPUTATION AND DATA PROCESSING BRANCH.

3. H. J. Juenemann
Acting Branch Chief

I. SUMMARY

The principal orientation of the Computation and Data Processing Branch during Fiscal Year 1967 has been focused on the conversion from Honeywell computer equipment to the newer IBM 360 line. This task has far exceeded all advance estimates as to the difficulty, complexity, man-hours and staff attention required. During the year, it has been necessary to modify the original goal--conversion of all work to a single 360/65 by September 1966--to the more realistic goal of converting that same workload to a multi-machine configuration (composed of two 360/50's and one 360/40) by August 1967. The original fixed price contract for conversion programming has been expanded from $375,000 to $484,000 and has been augmented by additional contractual support. At year end, the conversion is not yet completed with all Honeywell equipment remaining in use alongside of the new configuration. Discontinuation of the principal Honeywell equipment, two H-800 computers, is expected in early Fiscal Year 1968.

In spite of the large portion of its resources consumed by the conversion effort, the Computation and Data Processing Branch has, however, made significant contribution to NIH research by continuing to provide imaginative computational support under extremely difficult circumstances.

II. BRANCH PROGRAMS

A. Objectives

The Computation and Data Processing Branch provides a central facility for professional programming, computing, and automatic data processing services for NIH. It maintains operational liaison with research investigators, program officials and administrators in the effective use of its automatic computation and data processing services. The Branch performs all phases of centralized computation and data processing, using general purpose digital computers, data conversion equipment, other peripheral equipment and EAM equipment. Develops general purpose, user-oriented computer systems, routines and subroutines; coordinates the use of Branch automatic data processing equipment in peripheral locations; and exchanges technical knowledge and operating experience with other organizations engaged in computation and automatic data processing in support of biomedical research and related activities.

B. Current Programs

The Branch's major efforts have been and continue to be devoted to problems attendant on the equipment conversion. Retraining of personnel is nearing completion and the operational rationale for the new equipment is beginning to become sufficiently routine that the new equipment will soon be regarded by the programmer as a tool rather than as a curiosity.

A major effort is underway to expand and revise the library of standard general purpose computer programs which have wide utility at NIH. In addition, two major generalized computer languages of NIH origin, and specifically focused on the unique needs of NIH, have become operational on the new equipment. One of these, a tablemaking language, focused on the needs of the epidemiological statistician, has received wide acceptance. It is frequently used by programmers within the Branch who have need to rapidly respond to requests, but more importantly is being used directly by statisticians of the various NIH biometric branches to apply the power of computers to their immediate analytical needs in a relatively painless way.

Advice and assistance to NIH research programs have continued. Two relatively new programs have been introduced during Fiscal Year 1967. The Branch has assigned one of its staff members to an NIH field office in Panama, has arranged for a basic complement of ADP equipment, and has made a substantial contribution to the availability and usability of virological data collected by the NIAID field extension at the Panama site over the past seven years. The Computation and Data Processing Branch has also implemented its advice and assistance role in another important way. Ten centrally trained computer programmers have been temporarily or permanently relocated into specific laboratories at NIH as another way of bringing computational support more directly into the everyday environment of the research laboratory. The Branch's program of formal training of scientists in the Fortran language continues as yet another form of advice and assistance.

The Branch continues its service role as a nonbudgeted operation financed under the NIH service and supply fund on a fee-for-services basis. This method of financing insures that ADP services provided to NIH research and administrative functions are justifiable on a cost-versus-benefit basis. The growing involvement of central ADP services in the research and management functions of NIH is indicated by the fact that the net dollar value of the Branch's ADP efforts for Fiscal Year 1967 reached $3,445,905 with 500 individual tasks receiving attention.

C. Program Progress and Accomplishments

1. Analysis and Programming

The Branch function to provide computer programming capabilities in service to NIH research and management efforts and the necessary prerequisite functions of mathematical, statistical and data systems analysis,

occupy a primary role in the Branch's operation. Examples of specific efforts in these areas include the following:

a. The Branch undertook a cooperative effort with NHI to analyze the chemical energies involved in cardiac muscle activity during contraction in the presence of congestive heart failure. Initially, a discussion of the biochemical and physiological aspects of the study between the investigator and branch statistical programmers resulted in the decision that a multiple regression analysis should be performed on a normal control group in order to determine the most appropriate statistical model for the prediction of chemical energy utilization. Multiple linear regression by the stepwise analytic method reduced the model to three selected independent variables. The results for the control group were discussed with the investigator and were determined to be applicable in physiological terms. Subsequently, the experiment was repeated for two additional groups of muscles; one under the influence of norepinephrine and the other with induced congestive heart failure. Regression equations were obtained and, in conjunction with the investigator, a comparison was made of the regression equations obtained from the independent sets of data. It was concluded that of the three independent variables, only one might be significant. Partial correlations and partial regression coefficients will now be used in order to further differentiate between the three groups of muscle activity prior to any attempt to draw definitive conclusions.

b. A group of NIAMD investigators has for several years maintained, on an offsite computer, a set of computer programs providing tools for Simulation, Analysis and Model building (SAAM). This set of IBM 7094 Fortran II programs, consisting of approximately 150 subroutines, was developed over a number of years by Dr. Mones Berman and Mrs. M. F. Weiss, NIAMD. In order that SAAM would be available to NIH users on the IBM/360 system, the Fortran II version was converted to H-level Fortran IV for the 360. This was achieved in spite of the anomalities of the H-level Fortran compiler and the ambiguities and pecularities of the Fortran II version. As a result of this effort, a running program for SAAM is currently available for general use on the NIH IBM/360 and SAAM problems are running routinely.

c. In support of analytical work in DCRT's Laboratory of Applied Studies and the NIH Clinical Center's Clinical Chemistry Department, programs were developed to determine norms for fifteen routine chemical blood tests such as glucose, uric acid and calcium. Blood samples from 78 subjects for 12 consecutive weeks were analyzed. Computer programs were written to edit and tabularize the test results and to examine conditions such as (1) normality; (2) extreme results; (3) error measurements; (4) variations among subjects; (5) consistency of all blood chemistry results for all subjects; and (6) trends by the subject and groups of subjects. Current findings and computational aspects of the study are to be discussed with laboratory investigators and results of the study will be published in the medical literature.

d. In 1954 Dr. Harold Dorn, who was head of the Statistical Processing Section which was a forebearer of the present Computation and Data Processing Branch, undertook a study of a group of U. S. veterans to describe the relationship between tobacco use and mortality experience. This was a pioneering study in the application of machine methodology to large prospective epidemiological studies using rigorous statistical methodology. Two features of this study were unique; it was, in terms of the ADP technology at that time, an extremely large study made up of 293,658 subjects. Secondly, it was one of the first studies where precise definition of the population permitted identification and follow-up of both respondents and nonrespondents.

Beginning in January 1954, 198,834 V. A. policy holders responded to questionnaires on their smoking habits. In January 1967, a second questionnaire was mailed to nonrespondents, increasing the response rate to 85 percent. Mortality information was available from V. A. files. In subsequent years, Dr. Dorn continued to guide this study from positions in the Biometrics Branches of both NHI and NCI while this Branch has made the necessary arrangements with the V. A. to periodically update the mortality information and to provide the necessary statistical tabulations. During Fiscal Year 1967, the basic information file was again updated, but more importantly it was converted from the punched card technology, which was all that was available in 1954, to a computer based system geared to the long term follow-up of this pioneering study.

e. During the year the Pathology Retrieval System was converted from the H-800 to the IBM 360 and simultaneously its usability was improved. This system maintains a data file of autopsy, surgery and cytology information from patients' records. It also maintains a hierarchical dictionary of Systemized Nomenclature of Pathology (SNOP). The capability for rapid retrieval of relevant patients' records is provided through a multi-component query module in which searches are presented to the computer using logical combinations of the SNOP codes for topography, morphology, etiology and function, as well as patient characteristic and identifying information. The relevant subfile resulting from a query can be displayed to the pathologist through a variety of reports and summary tabulations. A natural language translator is being developed to enable the parsing of the pathology fact sheets which comprise the input information for the various files.

f. In support of the Cancer End Results Section of NCI, data for eleven tumor registries have been made available on the IBM 360 and computer programs have been prepared to accomplish the analysis. At the original request of WHO, eleven tumor registries were set up. Periodically data is sent to NIH from these (and other) registries with the most recent update of information for the eleven of current concern having been received in the third quarter of this fiscal year.

An edit program has been developed to check validity and consistency of the data and to identify special cases for further investigation. As a

result of this edit procedure, error listings and other information are sent back to the individual registries for correction of the files and a cancer master file is generated for use in analytical studies.

A system of survival analysis programs has been developed for the analysis of survival among the cancer patients. Parameters for a particular survival analysis can be provided to the program to govern the selection of cases and the cohort classification of selected data cases. Flexibility is provided to allow for minor variations in the calculations to be performed and to control the type of output desired. The survival analysis executed by these programs provides the numbers of alive, dead and lost cases in each cohort, in each interval of observation and calculates the observed, expected and relative probabilities of survival (both interval and cumulative interval) and the associated standard measures of error.

2. Computer Operations

The second most prominent function of the Branch is the operation of central computer equipment used by both Branch programmers and a growing number of "open shop" programmers. As has been noted, the equipment conversion is in process with both the Honeywell and IBM equipment being operated side-by-side. At the close of Fiscal Year 1967, the Branch was operating six computers dedicated to the general service requirements of NIH. Two 360/50's and one 360/40 represented the new generation of equipment while two H-800's and an H-200 represented the older generation.

At the beginning of the year, the Branch operated a 360/30 in support of the National Clearinghouse for Mental Health Information and the Cancer Chemotheraphy data processing systems. With the establishment of NIMH as a separate Bureau of PHS, the 360/30 was transferred to NIMH and substantial progress made to move the Cancer Chemotheraphy work to the 360/50's.

The second 360/50 was installed in May of 1967 and simultaneously the 360/40 was dedicated primarily to expedite the conversion of the administrative file processing activities of the Extramural Programs from the H-800's to the 360/40.

The computer operating functions of the Branch are faced with a constantly growing level of activity. This growth has never slowed since the first computer was installed at NIH in 1958. During the first quarter of Fiscal Year 1967, the monthly average of productive computer hours was 1568. The second quarter showed a monthly average of 1652 hours. During the third quarter, this figure rose to 1822 hours. (Fourth quarter figures are not yet available as of this writing.)

The peripheral computer support functions were also improved by installation of new plotting equipment. During the year, the Branch acquired a CalComp Model 663 Digital Incremental Plotter which is used for plotting results obtained from the IBM 360 computers. The system permits optional

switching between .005 and .0025 incremental step sizes to control speed and resolution of the plotted results. The system has greatly aided many NIH investigators in their biomedical research. For instance, the plotter is useful to visualize the periodicity of cardiac data as well as for visual display of the least-square fit of a theoretically derived biological model to observed laboratory data.

3. Computer Training for Scientists

The Branch's program for training scientists in basic computer technology continued throughout Fiscal Year 1967. This functional unit of the Branch provided 126 individual scientists with a basic course in computer programming and 1250 hours of "hands-on" experience on a small computer.

The training function of the Branch also is the focal point for providing advice and assistance to "open shop" scientific users, that is, to individual scientists on the NIH staff who prefer to write computer programs themselves rather than to have their problems programmed by a member of the CDPB staff. At the mid-point of Fiscal Year 1967, there were 134 authorized scientific "open shop" programmers, most of whom relied heavily on the staff of the Training Unit for assistance.

4. Machine Tabulation

The Machine Tabulation function contains the Electrical Accounting Machine (EAM) and "punched card" operations of the Branch. These "punched card" operations include three primary operations:

a. The keypunching operation of the Branch continued throughout Fiscal Year 1967 to operate beyond its capacity. This forced delays in service and necessitated occasional use of outside contracts. Approximately 300,000 cards are punched each month by Branch personnel and in addition, 60,000 cards per month are punched by outside contract. It is significant to note that throughout the fiscal year the cost has been consistently lower for inhouse keypunching as compared to contract keypunching. This cost benefit is maintained in spite of the fact that only the more simple jobs can be contracted out.

b. The EAM Unit, located at the Westwood Building, continues to provide specifically tailored punched card service to the Grants Management Offices located in the Westwood Building. This Unit has, throughout the year, maintained an outstanding level of service by consistently meeting tight deadlines in spite of continuous personnel shortages. The Westwood operation also includes a do-it-yourself facility for the exclusive use of four I/D Grants Management operations. Utilization of this Unit increased constantly during the year as more and more program managers availed themselves of the assistance available even from relatively unsophisticated equipment.

c. The on-campus EAM Unit at the central facility began during Fiscal

Year 1967 to undergo a reorientation. Much EAM processing previously had been related in some way to large on-going computer systems. These systems are gradually being redesigned to eliminate the EAM processing operations. Another major segment of the work of the EAM Unit is, for economic reasons, best processed on small, inexpensive equipment. Time on the H-200 (which is primarily a small but productive input/output slave to the H-800's) has been made available to the Machine Tabulation Section. Through the use of a few standard EAM simulator type programs, a significant portion of the EAM load is being reoriented to partial or complete H-200 processing, thus allowing release of five EAM machines.

Simultaneously, the Machine Tabulation Section is playing an evergrowing role in support of computer operations. During Fiscal Year 1967, access of programmers to EAM services needed in program preparation was emphasized with the designation of a separate group of keypunch operators for keypunching computer program decks and the development of an H-200 program which overcomes the incompatibility between the IBM 360 computer character set and IBM EAM equipment. In addition, seven keypunch machines, a tabulator and a sorter have been made available to programmers for "hands-on" use.

During the year the amount of laboratory-generated data arriving at the Branch in the form of punched paper tape continued to increase. All punched paper tape processing has now been moved to the H-200, allowing discontinuance of rental of the second Model 047 punched paper tape to card converter. All punched paper tape conversion is now made directly to or from computer magnetic tape.

5. Field Activities

In support of the Middle America Research Unit, NIAID, the Computation and Data Processing Branch activated during the Fiscal Year 1967 an ADP support unit in Panama. For approximately seven years, the staff of the Middle America Research Unit (MARU) has been involved in virological research in Central and South America. They have collected large amounts of data, including epidemiological and serological data, results of virus isolation studies and laboratory test results. These many scattered data became unmanageable under manual data handling techniques in terms of MARU's needs to (1) rapidly find results of interest; (2) correlate data from the same or different types of records; (3) keep an inventory of what data, results, sera, etc., are available and where they are located; and (4) selecting subsets of related data for intensive study.

Based on these needs, the Computation and Data Processing Branch placed a staff member on detail in Panama to establish an automated system for MARU. Initially, the "on-site" automation at MARU involved installation of EAM equipment and design of a new approach to data storage. Subsequently, the Computation and Data Processing Branch has also undertaken training of MARU staff and accomplishment of complex statistical calculations at its Bethesda location. Data available at MARU are

pre-processed there and forwarded to Bethesda for correlative statistical analysis. For example, we have undertaken a study attempting to relate ecology to prevalence of a specific parasitic infection. Sample data are being procured by MARU for every Central American country and comparable ecology data are being procured by the Computation and Data Processing Branch. Generalized data editing, tablemaking and data file computer programs have been written for this project. In addition to the presence of Computation and Data Processing Branch personnel at MARU, personnel from MARU will spend time working at Bethesda in order to gain a better understanding of the potential of large scale computers as well as to explore additional ways in which the central facility can provide support to MARU. Thirdly, the Computation and Data Processing Branch is collaborating with MARU personnel to write specific statistical computer programs that will be adaptable to a small scale computer available to MARU from another agency located in Panama.

6. Publications

Koudry, H. J.: Optimizing the systems design. Systems and Procedures Journal, pp. 26-29, September - October 1966.

Knott, G. D., Anderson, W. J., and McKay, J. A.: A table-making language. Proceedings of the IEEE, pp. 1779-1787, December 1966.

Lewis, M. S., Dubin, W., and Aandahl, A.: The physical properties of bovine corneal collagen. Experimental Eye Research, pp. 57-69, January 1967.

D. Problems

Recruitment

Recruitment of qualified professionals continues as the most serious Branch problem. Continuation of the Honeywell equipment joined with the expansion of the number of IBM computers operated by the Branch and coupled with normal turnover in the highly competitive market have generated an acute need for computer operations personnel. Continual expansion of the number of type and complexity of NIH research programs demanding computational support have generated a constantly acute need for experienced scientific programming personnel. A temporary freeze to the September 30, 1966, employment level seriously harmed recruiting. Since that freeze has been removed, cumbersome recruiting and certification procedures applied to an extremely competitive market have prevented an adequate solution of this problem. The nationwide shortage of qualified programmers has hampered the development of an adequate central computer facility staffing pattern and has seriously inhibited the Branch's internal operation as well as its program of placing experienced computer professionals directly in the laboratories of NIH.

July 1, 1966 through June 30, 1967

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Laboratory Activities

1. DCRT-3
Serial Number

2. LABORATORY OF APPLIED STUDIES

3. Eugene K. Harris, Ph.D.
Chief

I. GENERAL SUMMARY

The major effort of this newly-formed Laboratory has been to identify and explore promising areas of biomedical research where the imaginative application of computer technology can make a unique contribution. The Laboratory is particularly interested in applications of mathematical statistics, applied mathematics, and computer science to clinical research at NIH. Included within this definition of clinical research are I/D experimental laboratory studies with clearly definable implications for the understanding and treatment of disease. Through collaborative efforts with I/D scientists and clinicians, a number of Laboratory research projects have been defined and are now underway in areas of cardiology, neurology, and metabolic studies. In addition, the Laboratory has initiated several independently-generated research efforts in computer-related areas of mathematical statistics (e.g., random sampling, multivariate analysis). The Laboratory is also directing the development of linked computer programs for statistical analysis of time-dependent processes and offers formal training for scientists interested in the theory of time-series analysis.

II. CURRENT LABORATORY PROGRAMS

A. Objectives

Current programs of this Laboratory may be described in terms of the following goals:

1. Research, construction, testing, and refining of mathematical statistical or biophysical models of physiological and general biological processes of medical significance.

2. Initiation of sound collaborative research involving the application of mathematical statistical theory and related computer technology to problems of importance in clinical medicine. In support of research projects, specialized computer programs must be written (a) to test biomathematical models, (b) to simulate biological processes too complex for explicit mathematical solution, (c) to provide mathematical statistical analysis of multivariate laboratory and clinical data, often collected continuously in analog form and requiring analog-digital conversion.

B. Summary of Current Projects

1. Collaborative Studies with Surgery Branch, NHI, and Surgical Neurology Branch, NINDB.

Biomathematical studies have been initiated by this Laboratory during this fiscal year in collaboration with the Surgery Branch of the National Heart Institute and the Surgical Neurology Branch of the National Institute of Neurological Diseases and Blindness (Individual Project Reports: DCRT 3.6 - 3.8). These studies concern the construction and testing of mathematical models and the application of advanced methods of statistical analysis in such areas as (a) computer simulation of arrhythmias and analysis of EKG features in arrhythmic patients (DCRT 3.6); (b) the role of each cerebral hemisphere in defining the relationships between the EEG, EKG, palm and rectal temperatures in neurosurgical patients and in "split-brain" chimpanzees (following mesial cerebral incision as described in a recent paper by Baldwin, et al., NINDB) (DCRT 3.7). In the study of cardiac arrhythmias the Laboratory is interested in two general areas, (1) simulation of the EKG as a function of membrane potential changes in atrial and Purkinje fibers along the cardiac conduction pathway (DCRT 3.12); (2) exploration of statistical theory of random events to determine whether certain statistical characteristics (e.g., so-called "intensity" functions related to the probability of occurrence of an event at a given time since the last event) can be associated with specific types of arrhythmias for use in diagnosis and evaluation of treatment.

Studies such as these require the acquisition in continuous time of large volumes of multivariate data and reducing these data to digital as well as graphical form prior to mathematical model building and analysis. Experience has shown that improved methods of data recording need to be introduced into the laboratories or surgical suites while, concurrently, comprehensive computer programming systems are being prepared. For example, current use of multiple strip charts to record continuous data during studies of "split-brain" chimpanzees should be replaced by analog tape which will, in turn, require programs for analog to digital conversion of multiplexed tape channels.

Again, in the Heart Surgical Suite, the Laboratory recognized the need for versatile editing equipment to permit careful selection of recorded data for specific mathematical analyses. The tape system currently available in the surgical suite was found to be inadequate for editing purposes. In cooperation with the Computer Systems Laboratory, DCRT, the required equipment was specified, purchased, and installed in the Clinical Center.

To support these and other clinical studies, an effort was launched late in the fiscal year to produce a much-needed package of linked time series analysis programs for the IBM 360 (DCRT 3.9). These programs included facilities for autocovariance and autocorrelation, Fourier transforms, power spectrum analysis, and related techniques (trend removal, pre-whitening, etc.). Such programs are vital to all but the most superficial mathematical statistical analyses of the multivariate, time-dependent data collected during these studies.

As part of the same program of collaborative research with the surgical branches of NHI and NINDB, this fiscal year saw the completion of a system of linked computer programs to analyze continuously recorded measurements of ventricular and aortic pressures, blood flow, and the corresponding EKG (DCRT 3.5). Although this analysis will be off-line on the central IBM model 360 computer, it will permit much more rapid evaluation of the effects of surgical procedures, including rates of change of heart work and power and energy losses across the aortic valve. In addition, these programs will greatly facilitate testing currently available mathematical models of heart mechanics to determine their usefulness in distinguishing different kinds of heart disease. Investigations of normal variations in cardiovascular parameters from beat to beat will also use these programs.

In short, projects initiated during this fiscal year by this Laboratory in cooperation with NIH research surgical programs have had to concentrate on providing a good foundation for later mathematical statistical or biophysical analysis, model-building and testing. This essential phase, involving the establishment of data recording and editing facilities and the development of computer programming systems will continue into FY 1968. However, the coming fiscal year should see the first results of biomathematical research based on the improved methods of data management introduced this year.

2. Recovery and Classification of Bioelectric Signals.

One research project of this Laboratory, in collaboration with the Laboratory of Neurophysiology, NINDB, concerns a mathematical statistical investigation of the operation of a new adaptive filter technique for the recovery of neuroelectric signals of either fixed or variable latency (DCRT 3.16). The method uses iterative correlation-averaging procedures, converging to a recovered waveform in a template record. After each cycle, the frequency distribution of deviations from convergence are displayed as well as the maximum cross-correlations observed between the template and replicate sample records. It appears from study of both artificial and bioelectric signals that relatively simple statistical analyses of this output permit the separation of nonrandom from random noise and assessment of that portion of the former associated with the definition of the signal itself.

More broadly, the technique of adaptive filtering has application to pattern recognition in general and has been found useful in classification of electrocardiograms and, possibly, electroencephalograms. Previous adaptive filtering techniques have used matched filters. The present correlation-averaging technique seems better suited to establish a norm for each individual based on his own responses at intervals over past time. The current program now on a LINC computer, will be programmed for the IBM 360 early in FY 1968. Coupled with the newly-developed capability for computer handling of EKG waveforms, this Laboratory should be in a position during FY 1968 to begin more intensive study of the potential of various adaptive filter techniques in the classification of bioelectric signals.

3. Mathematical Statistical and Related Computer-Based Projects

Other research projects (DCRT 3.10 - 3.12) of this Laboratory are essentially mathematical statistical studies in areas where data acquisition problems have already been solved and large portions of the computer programming systems are in operation.

In one such project, oriented to the efficient use of computers in biological model building, a generalized, efficient program for fitting non-linear models has been designed. This program, built as a modular set of subroutines, will permit remote users to evaluate any mathematical model expressible in Fortran notation. The new subroutines are now being evaluated in studies of sodium and calcium kinetics and the response of the parathyroid to changes in serum calcium levels. These studies are in collaboration with scientists from NCI, NIMH, and NHI.

Statistical analysis of clinical laboratory results, a project begun in fiscal year 1966 in collaboration with the Clinical Pathology Department, Clinical Center, has been actively pursued during this fiscal year (DCRT 3.4). Attention this year has been focused on the evaluation of sources of variation in blood chemistry results from apparently healthy individuals of different age, race and sex groups. Some fifteen blood tests are currently under study.

A "components of variance" model has been applied to separate variation in results from each individual into two independent parts, one reflecting inherent physiological variability, the other arising from fluctuations which occur when the same blood specimen is sampled and examined repeatedly under routine laboratory conditions using the same analytical method. Daily analyses of frozen pooled sera provide data from which the analytical variation in each blood test can be ascertained. Detailed studies of the relative importance of different sources of variation in blood chemistry measurements among healthy individuals are surprisingly rare. This study, although limited with respect to the depth of sampling among the multitude of socio-economic, dietary and other groupings, should be of considerable interest to physicians concerned with normal ranges, and a valuable guide to future investigations.

In another project, undertaken at the request of the Associate Director, Clinical Center, an algorithm has been developed to computerize the scheduling of interviews between NIH scientists and candidates for research and clinical associate positions (DCRT 3.15). The automated system, presently working in parallel with the current manual scheduling procedures, is expected to replace the latter by January, 1968. The new system will substantially ease the difficult task of arranging hundreds of individual candidate-scientist interviews which must be done within the space of a few weeks each spring.

C. Other Activities; Future Plans for Training; Organizational Developments

Laboratory staff members served on the faculty of the Foundation for Advanced Education in the Sciences, teaching two courses: 1) Elementary

Probability for the Biological Sciences, and 2) Time Series Analysis. A short series of lectures was held on the uses of decision theory in management.

One long-term purpose of the Laboratory concerns the training of NIH biomedical scientists and clinicians in the understanding and use of biomathematical models, statistical methodology and associated computer programs. This training will be implemented in several ways: 1) selection of qualified, interested physicians, preferably after two or three years of research or residency experience, to join the Laboratory for several years during which intensive training in computer programming, mathematical statistics and the management of quantitative research data (including the use of modern data acquisition devices) will be followed by full-time research in some area of interest to both the Laboratory and the individual, generally in collaboration with one or more I/D research workers outside DCRT; 2) the development of short daytime courses in the management and analysis of clinical and clinically-related data with particular reference to the construction and testing of mathematical models and the statistical analysis of time series; 3) the preparation and distribution of complete descriptions of specialized computer programming systems developed under the direction of this Laboratory. This will be followed up where necessary by informal seminars to ensure user familiarity with these programs.

As the Laboratory's research interests began to crystallize during the latter half of FY 1967, organization into sections seemed appropriate. At the end of this year, two sections were established: Statistics and Quantitative Biology Section (James E. Mosimann, Chief), and Biomedical Studies Section (Eugene K. Harris, Acting Chief).

Note: Two research projects identified in the DCRT Annual Report for Fiscal Year 1966, Numbers 3.2 and 3.3, are no longer being pursued by this Laboratory. Number 3.2, previously called Automated Sequencing Laboratory has been transferred to the Computer Systems Laboratory, Number 4.2 (Real-Time Computer Support for Biochemistry). Number 3.3 (Basic Studies of Viscous Effects on Flow Mechanisms of Biological Fluids in Capillaries and Other Microscopic Structures), has been discontinued following departure of the principal investigator.

Serial No. 3.1
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Mathematical Problems of Sequence Determination of Nucleic Acids and Proteins

Previous Serial Number: 3.1

Principal Investigator: James E. Mosimann, Ph.D.

Other Investigator: Jay Vinton

Cooperating Units: Computer Systems Laboratory, DCRT

Man Years (computed for the 12 month period)

Total:	0.2
Professional:	0.2
Other:	0.0

Project Description:

Objectives:

1. To provide mathematical definitions of classes of protein and nucleic acid sequences which cannot be determined using complete digest methods.

2. To provide a probability framework for evaluating the probability that protein and nucleic acid sequences of certain classes are solvable, using partial digest methods.

3. To provide mathematical support necessary for testing computer programs designed to reconstruct nucleic acid and protein sequences from fragment data.

Methods Employed:

The primary mathematical tools are (1) algebra, specifically, the algebra of free monoids; (2) graph theory; and (3) stochastic processes, Markov chains.

Major Findings:

There are broad classes of sequences which cannot be determined by complete digests with monobase specific enzymes. The theoretical studies clearly indicate the need for partial digest fragment data in determining sequences of nucleic acids.

Significance to Biomedical Research:

Determination of the primary structure of proteins and nucleic acids is of extreme importance in studies of the genetic code, as well as molecular bases of life.

Current Status:

Work is underway to establish necessary and sufficient conditions for a sequence to be solvable.

Honors and Awards: None

Publications:

These publications were listed last year as being in press and are listed this year to provide the full citation.

1. Reconstruction of protein and nucleic acid sequences IV. The algebra of free monoids and the fragmentation stratagem. Mosimann, James E. (with M.B. Shapiro, C. R. Merrill, D. F. Bradley, and J. E. Vinton). The Bulletin of Mathematical Biophysics, Vol. 28, 1966, p. 235-260.

2. Reconstruction of protein and nucleic acid sequences V. Computer-simulated tests of various tactics for reconstructing the sequences of transfer ribonucleic acids. Mosimann, James E. (with C. R. Merrill, M. B. Shapiro, D. F. Bradley, and J. E. Vinton). Biopolymers, Vol. 4, 1966, p. 723-735.

Serial No. 3.4
1. Lab. of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Statistical Research in Clinical Pathology

Previous Serial Number: 3.4

Principal Investigator: Eugene K. Harris, Ph.D.

Other Investigators: Paul Kanofsky, Ph.D., NCI; George Shakarji, DCRT, CDPB

Cooperating Units: Clinical Pathology Department, Clinical Center

Man Years (computed for the 12 month period)
Total: .85
Professional: .75
Other: .10

Project Description:

Objectives:

To explore, through careful application of statistical methodology, relationships inherent in large volumes of results of chemical analyses routinely performed on serial blood samples from normal volunteers and patients at the Clinical Center.

Methods Employed:

At this time, blood chemistries from 79 normal volunteers, in ten groups of seven or eight individuals per group, are stored in easily retrievable form on digital magnetic tape. These data consist of concentrations (milli-equivalents per liter) of at least 15 elements and compounds (including some enzymes) measured in each of 10-12 weekly blood samples from each subject. The total ten-group data base covers a 2 1/2 year period of study.

Similarly stored patient blood data are available from approximately 2500 Clinical Center admissions during a 9-month period in 1964-65. Demographic and diagnostic information on these patients is also available for related study. Serial blood data for each patient is less abundant than for normal volunteers and not regularly spaced, but does permit some time-wise analyses relative to effects of therapy or other changes of state.

A separate component of variance analysis has been applied by computer program to the weekly concentrations for each normal volunteer (79 subjects, 15 blood tests, each in duplicate, 10-12 samples per subject -- a total of approximately 25,000 observations, 300-350 per subject). The availability of daily measurements on pooled frozen serum allowed separate estimation of analytical error, so that weekly variation in a subject's results can be apportioned to (a) inherent physiological variability, (b) fluctuations in routine laboratory analysis of replicate samples. Tests of normality using Wilk's criterion have been applied to the mean weekly results from each subject.

Major Findings:

At this point in time, results of the variance component analysis are not yet complete, so that the extent of variability in normal controls attributable to inherent physiological fluctuation as contrasted with analytical, laboratory variation cannot now be reported. Frequency distributions of these variance components over all subjects, and by age-sex-race grouping will soon be available, however. It does appear that within each individual weekly means of duplicate analyses are approximately normally distributed.

Significance to Biomedical Research:

These studies will (1) add substantially to existing information on physiological variation of blood constituents within a single healthy individual over time and among normal individuals within a given age, sex, and race group, and (2) provide new knowledge on currently achievable analytical precision, and (3) provide an opportunity to test potentially useful mathematical models of time-dependent changes in blood constituents of patients during the course of therapy.

Proposed Course of Project:

Following completion of analysis of data from normal volunteers, attention will be centered on the patient file, using distributions of blood tests in normal volunteers to aid in isolating extreme ranges in each variable among patients. Correlations among extremes and relationships to diagnosis and to demographic variables will be examined, and relevant physiological hypotheses will be tested. Finally, the variation in each test over time will be studied in relation to early and final diagnoses and course of therapy. At this point, more sophisticated mathematical statistical methods, drawn from the theory of stochastic processes, will need to be developed to describe dynamic changes in test results during hospitalization.

Honors and Awards: None

Publications: None

Serial No. 3.5
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Heart Surgical Suite Project

Previous Serial Number: None

Principal Investigator: David B. Gilbert, M.D.

Other Investigators: Felicity Callahan, CDPB
James Standish, CDPB

Cooperating Units: Surgery Branch, NHI (cooperation includes data files from Heart Surgical Suite; no funds are provided)

Man Years (computed for the 12 month period)
Total: 1.1
Professional: 1.1
Other: 0.0

Project Description:

Objectives:

The goals of the Heart Surgical Suite Project are to provide accurate, well-organized mathematical analyses of massive quantities of physiological data obtained from humans during surgery.

Methods Employed:

Data acquired in analog form on magnetic tape during surgery are edited off-line on equipment designed by the Division. The data are then converted to digital form on the CDC 3100. The digital tape is analyzed off-line on the IBM 360 system utilizing Fortran and assembly language programs. The parameters presently studied include two pressures, the flow between these two pressures, and the electrocardiogram. Classical analyses of this data include the maxima and minima of flow and pressure, rates of flow and pressure change, stroke volume, stroke work and power. Experimental analysis includes the energy loss across a diseased valve expressed as valvular efficiency and the distribution of the fractional energy loss throughout a cycle. Reverse flow in the aorta is also calculated.

The technique used may be described as follows. An EKG is used as a marker for the onset of a cardiac cycle. This is accomplished by two subroutines that determine the baseline and subsequently recognize a QRS or premature complex and distinguish it from high frequency noise or a T-wave. Pressure and flow curves are smoothed by Fourier analysis and resynthesis from the coefficients of a given number of harmonics. The coefficients are further utilized to calculate the derivatives and integrals of the curve as well as for computing the impedence as a function of each harmonic. Embedded in the routine is a least squares line and parabola fitting program that, in addition to the Fourier analysis, may be used independently.

Major Findings:

The Fortran programs required for the above off-line analyses are nearly complete. A system for editing and acquiring analog data is in the final stages of completion. The feasibility of analyzing this quantity of data on-line or during surgery seems remote but will be better judged after operation of the program off-line on the 360 system.

A theoretical model that will quantitate energy loss across a stenotic valve has been developed and may prove valuable in describing the course of restrictive valvular lesions.

Significance to Biomedical Research:

Considerable quantities of analog data are available in the area of cardiac research, particularly in the measurement of pressure, flow and the electrocardiogram. Due to the complexity of the calculations, relatively few portions of this data are analyzed extensively. Techniques developed here should allow more thorough analysis of that data including the variability of one individual's parameters.

Proposed Course of Project:

By the end of FY 1967, the software package should be operational. This will allow preliminary analysis of data already collected from patients with idiopathic subaortic and calcific stenosis. Thereafter the package will be available to the NIH campus and scientific community at large.

After initial analysis predictions may then be made as to the volume and quality of data as well as the language format that would be required for on-line processing. Applications to cardiac catheterization data will then be studied.

Honors and Awards: None

Publications: None

Serial No. 3.6
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Computer Modeling of Arrhythmias

Previous Serial Number: None

Principal Investigators: David B. Gilbert, M.D.
Eugene K. Harris, Ph.D.

Other Investigators: None

Cooperating Units: Surgery Branch, NHI (to supply data in the form of magnetic tapes)

Man Years (computed for the 12 month period)
Total: 0.4
Professional: 0.4
Other: 0.0

Project Description:

Objectives:

The immediate goal of this project is to simulate the sequential changes expected in cardiac conduction during rapid stimulation using the digital computer. Ultimately, the simulation of sequential electrical activity should aid in pattern recognition studies for classification of arrhythmias.

Methods Employed:

Utilizing known facts about transmembrane potential and thresholds of response for each of the pathways of cardiac conduction, a digital computer simulation of cardiac conduction has been programmed. The response of each nerve fiber is determined by its previous response and the input impulse. For a given number of cardiac cycles, the model will simulate and plot the action potential of the sinoatrial node pacemaker, and atrial fiber conducting an impulse to the atrioventricular node and the effect of the A-V node upon a ventricular fiber.

Major Findings:

Progress has been satisfactorily completed to date to predict normal conduction responses. Work has begun upon adding permutations to portions of the conduction system to test the robustness of the model.

Significance to Biomedical Research:

A general need is recognized for the adequate training of physicians to interpret cardiac arrhythmias. The proposed model graphically depicts each step of cardiac conduction as well as the expected time at which one would expect the atria or ventricle to fire. More specifically, the model should aid to our knowledge about the ventricular response to atrial fibrillation and how that response might be altered.

Proposed Course of Project:

The model is to be tested systematically for its ability to predict known changes recognized on the EKG. Data from individuals in known atrial fibrillation will be collected and the resultant frequency distributions of the R-R interval will be compared to the digital computer simulation. The effects of respiration upon that distribution can be then studied both from a theoretical and experimental point of view.

Honors and Awards: None

Publications: None

Serial No. 3.7
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Application of Advanced Mathematical Techniques in Surgical Neurology

Previous Serial Number: None

Principal Investigator: Raymond Mejia

Other Investigators: None

Cooperating Units: Surgical Neurology Branch, NINDB

Man Years (computed for the 12 month period)
Total: 0.8
Professional: 0.3
Other (including computer time): 0.5

Project Description:

Objectives:

The objective is to study physiological parameters available during surgery through the application of advanced digital signal processing techniques. Studies will also be carried out through the analysis of patient data obtained in the Neurology recording facility. Mathematical and mathematical statistical methods for the analysis of gross neurological data will be developed and applied.

Current and Proposed Methods:

A study of contralateral and ipsilateral activity has been initiated using data from animals and patients. Variables include EEG, EKG, palm and rectal temperatures measured during surgery as well as pre- and post-operatively. Currently, multiple strip chart and analog recording equipment is being used to gather this information. Improved procedures for collection and digitization will be established, including recoding of data on analog tape and programming for analog to digital conversion of multiplexed channels. Initially, a -> d programs will be written for the LINC computer and digital data put in a form directly acceptable to the IBM 360.

Major Findings:

At this time, effort is concentrated on improving methods of data acquisition and preparation for digital computer processing. Results of mathematical analysis can be expected during Fiscal Year 1968.

Honors and Awards: None

Publications: None

Serial No. 3.8
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: A Mathematical Model to Study Cerebral Concussion and Related Brain Trauma

Previous Serial Number: None

Principal Investigator: Herbert Reff

Other Investigator: Ayub K. Ommaya, M.D.

Cooperating Units: Surgical Neurology Branch, NINDB

Man Years (computed for the 12 month period)
Total: 1.5
Professional: 1.5
Other: 0.0

Project Description:

Objectives:

A determination of those physical parameters associated with cerebral concussion and related brain trauma. The determination of these parameters based upon a mathematical model can suggest directions which could be taken to prevent or decrease the severity of such injuries and lead to a better understanding of concussion and brain trauma.

Methods Employed:

1. Experimental cerebral concussion is produced by the application of a blow to the occipital region of the head. (Rhesus monkeys are currently being subjected to these experiments.)

2. Data are currently obtained from the digitization of head displacements, recorded by high-speed cinematography, in a plane perpendicular to the axis of the camera. It is also planned to encompass the movement of the brain with respect to the skull through the utilization of ultra high-speed X-ray analysis of the displacement of iso-density radio-opaque markers in the subject's brain.

3. The data are then smoothed and mathematical functions will be derived to represent the equations of motion of the system.

Major Findings:

It has been indicated that acceleration and in particular angular accelerations which are developed by the head as well as by the brain with respect to the head are quite important in inducing brain concussion. The brain injuries that appear most frequently consist mainly of cerebral hemorrhages which occur at the surface of the cortex. This indicates that the development of shear stresses during these experiments can, if allowed to surpass certain levels, cause serious damage within the system.

Proposed Course of Project:

It is suggested to continue these studies along the current lines to determine precisely the acceleration levels which must be attained to induce cerebral concussion as well as the values of shear stress required to produce tearing of blood vessels. Once these determinations have been made, methods may be devised to prevent accelerations from exceeding tolerable levels.

Honors and Awards: None

Publications: None

Serial No. 3.9
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Time Series Analysis Package

Previous Serial Number: None

Principal Investigator: Raymond Mejia

Other Investigators: None

Cooperating Units: Computation and Data Processing Branch, DCRT

Man Years (computed for the 12 month period)
Total: 1.0
Professional: 0.3
Other (including computer time): 0.7

Project Description:

Objectives:

There is a need at NIH for a library of computer programs to aid the analysis of time-dependent data. Benefits from such a set of programs will accrue to the researcher who has collected small amounts of data over a relatively long period of time as well as to the project in which real-time data analysis is required to control physical and chemical processes.

At this time it is desirable to make available a wide class of time series techniques to users of the IBM 360/50 computer at NIH. These programs will be made available to remote users of the system when remote terminals are available. In conjunction, the implementation of a limited library facility through a teletypewriter terminal remote to a GE 235 computer will be undertaken. The purpose is to make programs available remotely through the means currently available with a minimum of duplication and to determine the desirability of a conversational Fortran language for use by non-programmer scientists.

Methods and Plans

The numerical techniques for most of the analysis programs to be included in this library are well known, and many of the programs are

presently implemented in some form. Hence, most of the work required will be in documentation, conversion to Fortran IV and adaptation to the multiprogramming monitor. Executive routines will be required to facilitate linkage of programs and passing of data sets from one program to another. All programs will require complete documentation, adhering to NIH program library conventions, which are similar to the SHARE standard. In addition, symbolic decks of programs will include comments to document handling and exchange procedures as well as updating or modifying programs.

Executive routines will be required to permit selection and sequencing of operations and to provide for data handling within the set of programs specified.

Programs for "raw" data processing will include such functions as:

a. acquisition of single channel and multiplexed data
b. smoothing
c. text of a trend
d. detrending
e. prewhitening

Programs for correlation analysis will include:

a. autocorrelation
b. cross-correlation
c. higher order correlations
d. smoothing or lag windows

Programs for frequency analysis will include:

a. harmonic analysis
b. spectral analysis
c. cross-spectral analysis
d. calculation of coherency and transfer functions
e. spectral windows

In addition to providing printed output at remote terminals and at batched terminals (to be forwarded to the user), plots of many of these quantities will be desired. Hence, routines to generate tapes for the NIH-DCRT plotter must be available to remote users.

Test programs for all routines and test data generators for investigations in time series analysis as well as simulation will include generators of periodic, aperiodic and random data.

Honors and Awards: None

Publications: None

Serial No. 3.10
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Biological Model Evaluator Studies

Previous Serial Number: None

Principal Investigator: John E. Fletcher

Other Investigator: Richard Shrager

Cooperating Units: CDPB, Unit #5

Man Years (computed for the 12 month period)
Total: 0.5
Professional: 0.5
Other: 0.0

Project Description:

Objectives:

The object of the Biological Model Evaluator Studies is twofold. The primary object is to collect and develop those computer-oriented mathematical and statistical tools most applicable to the testing of biomathematic models. These tools are to be assimilated into user oriented program packages which provide NIH with a generalized biological model evaluator capability. Secondly, this project acts collaboratively in developing and supplementing mathematical-biological model building in other Institutes at NIH. In so doing, the requirements of biomedically oriented investigators for model evaluation are kept current and the capability of the evaluator programs can be updated accordingly.

Methods Employed:

1. The central concept of this project is model fitting by means of mathematical techniques. Two systems are currently operating. The first is the SAAM (Simulation Analysis and Modeling) system developed by the mathematical research branch of NIAMD, and the second is a nonlinear curve fitting program developed by unit #5, CDPB, DCRT. The SAAM system is essentially a closed-package system of computer programs built around the concept of compartmental analysis and its

associated techniques. This system cannot be used in connection with other programs, and special modifications are necessary for all but a small class of well-defined mathematical models. The SAAM system has operated for a number of years at the National Bureau of Standards and other installations and is a proven operating tool. It has the advantage of being a large complete program, for which only data format need be expressed.

The second system is a class of subroutines (NIH 22-23) designed to do nonlinear least squares curve fitting for a general class of mathematical models. To use this system, a small Fortran program must be written defining the proposed model. The programming system accepts any model expressible in Fortran notation. In addition, it may be used as a subsystem in a larger user-oriented program. These subroutines have just become operational and are still in the advanced testing stage. Upon completion of satisfactory testing, these newer, more versatile techniques will be documented and distributed for NIH use. Further development of this project will permit the inclusion of data handling subroutines or numerical analysis subroutines which may enrich its utility as a biomedical tool.

2. Collaborative efforts are directed to the following projects:

NIH - NCI: Metabolism Branch; Dr. James Phang. The Quantitative Determination of Calcium Kinetics and Bone Synthesis in Normal Subjects.

NIH - NIMH - LCS: Unit on Psychosomatics; Dr. Leslie Baer. The Quantitative Determination of Sodium Kinetics in Normal and Diseased Patients.

NIH - NHI: Section on Polypeptide Hormones, Laboratory of Molecular Disease; Dr. John Potts. Parathyroid response to changes in serum Ca level

These projects are similar in nature and are discussed in the reports of the respective Institutes. The participation in these efforts has involved basic formulation of an appropriate model for clinically acquired experimental data, data processing of the data and model, and interpretation of results. Refinements of data collection, experimental methods, and biological models are often suggested as a consequence of mathematical results obtained through the computer program.

Proposed Course of the Project:

Continuing efforts will (1) evaluate the two alternate model-testing programs mentioned above; (2) catalog supplementary subroutines which increase the utility of the model evaluator system; (3) begin extensive documentation and educational materials for training prospective users in these mathematical-computer tools.

Honors and Awards: None

Publications: None

Serial No. 3.11
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Computer Simulation Methods and Related Statistical Methodology

Previous Serial Number: None

Principal Investigator: Robert J. Connor, Ph.D.

Other Investigators: Robert E. Hackman
James E. Mosimann

Cooperating Units: None

Man Years (computed for the 12 month period)
Total: 2.25
Professional: 2.00
Other: 0.25

Project Description:

Objectives:

To provide NIH with special computer-related statistical models.

Major Findings:

1. A pseudo-random number generator has been tested and implemented. The generator is one suggested by Marsaglia (A General Method for Producing Random Variables in a Computer, Proceedings of the 1966 Fall Joint Computer Conference). The generator has been programmed and is available on system 360. It satisfies uniformity to the third digit, but breaks down at the fourth. Some form of serial dependency exists in the series, although its serial correlation structure is acceptable.

2. An algorithm to provide a rapid means of determining equiprobable intervals of a probability density function has developed. This has resulted in a tenfold (from 50 minutes to 5 minutes) reduction in computer time over previously used algorithms. The currently used algorithm has been used to subdivide the range of normal and chi-square random variables into 1000 equiprobable intervals.

3. A statistical model has been developed to permit the analysis of regression data for equations with the same slope but possibly different intercepts. The model is relevant to the combination of several sets of experimental data, where the results of each set are measured from a different baseline. Such data is common in biomedical research. The results are programmed on system 360.

Honors and Awards: None

Publications: None

Serial No. 3.12
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: EKG Modeling and Analysis

Previous Serial Number: None

Principal Investigator: David B. Gilbert, M.D.

Other Investigators: None

Cooperating Units: None

Man Years (computed for the 12 month period)
Total: 0.3
Professional: 0.3
Other: 0.0

Project Description:

Objectives:

The goal of this modeling project is the development of a mathematical model to predict the character of a scalar EKG as a function of the transmembrane potential of atrial and Purkinje fibers.

Methods Employed:

The transmembrane potential of a Purkinje fiber is described as a function of sodium and potassium conductance. A series of these potentials are summed as a function of the time at which they might be expected to fire. The present model assumes that the EKG represents the time derivative of that summed transmembrane potential. A digital computer is utilized for all calculations and subsequent inscription of the predicted EKG.

Major Findings:

The model presently coarsely predicts the changes expected for hypertrophy, premature contraction and ischemia. It inaccurately inverts the "T-wave" in all instances. The model has led to the development of a preliminary hypothesis that would state that electrographic hypertrophy is a result of increased conduction velocity.

Significance to Biomedical Research:

The EKG has been a most useful clinical empirical tool. Principal mathematical and computer techniques have centered about the organization and optimization of this data in 3-space. Relatively little emphasis has been placed upon the origin of the biological signal; i.e., the transmembrane potential. A suitable model is needed that will predict electrocardiographic waveform as a function of the variables producing (as opposed to modifying) that waveform.

Proposed Course of Project:

Should the mathematical model be shown to be robust, it will allow alterations in sodium or potassium conductance and predict variations in the EKG waveform. This should develop a number of hypotheses in regard to the biochemical cause of a given EKG waveform. Such hypotheses may then be systematically evaluated through computer simulation based on parameter changes in the model.

Honors and Awards: None

Publications: None

Serial No. 3.13
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: The Statistical Analysis of Percentage or Proportional Data

Previous Serial Number: None

Principal Investigator: James E. Mosimann, Ph.D.

Other Investigators: Robert J. Connor, Ph.D.,; Robert E. Hackman

Cooperating Units: None

Man Years (computed for the 12 month period)
Total: 1.25
Professional: 1.00
Other: 0.25

Project Description:

Objectives and Methods:

1. To provide mathematical and statistical models to permit the analysis of data which are proportioned, e.g., per cent serum protein constituents of blood.

2. To provide mathematical support for biological studies of growth and differential growth. The primary mathematical methods involved are those of probability and mathematical statistics along the lines of the multivariate beta or Dirichlet model (Mosimann, 1962, 1963).

Major Findings:

A general class of multivariate distribution whose variables are proportions has been developed. The general class describes a competitive situation in which one particular proportion has a dominant role to some other proportion or set of proportions. This means, for example, that one variable assumes its share of a resource without reference to the amount the second variable then subsequently assumes. The models include as a special case the Dirichlet model. Hence, they provide alternative hypotheses to the simpler non-competitive hypothesis associated with the Dirichlet model. In analyzing data on differential growth published by Mosimann, the models have proved

to be of biological use in identifying which of several variables shows growth dominant to other variables. This phase of the project is being written for publication at the present time. The possibility of collaboration with Dr. A. J. Coulombre of the Institute of Child Health and Human Development is also being discussed.

References:

1962. On the compound multinomial distribution, the multivariate beta distribution and correlations among proportions. Biometrika. Vol. 49, No. 1-2, p. 65-82.

1963. On the compound negative multinomial distribution and correlations among inversely sampled pollen counts. Biometrika. p. 47-54.

Honors and Awards: None

Publications: None

Serial No. 3.14
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Molecular Model Building Using Theoretically and Empirically Determined Intra- and Intermolecular Potential Functions

Previous Serial Number: None

Principal Investigator: William P. Minicozzi, M.D.

Other Investigators: Dan Bradley, M.D.; Howard Nash, M.D.

Cooperating Units: Laboratory of Neurochemistry, Section on Physical Chemistry, NIMH

Man Years (computed for the 12 month period)
Total: 0.8
Professional: 0.8
Other: 0.0

Project Description:

Objectives:

To develop a set of potential functions which can be used to calculate the energy of various molecular configurations for any molecular system.

Methods Employed:

X-ray crystallographic and electron scattering data are used to determine bond lengths and fixed bond angles for the molecular system under consideration. Then a linear combination of atomic orbitals is used to calculate charge distributions in the molecule under consideration to the various atoms consistent with the experimentally determined dipole moment of the molecule. An atom-atom interaction energy is the sum of four components in our model:

1. Electronic repulsion due to overlap of orbitals.

2. Electron correlation energy arising from the induced instantaneous dipoles.

3. Electrostatic energy, calculated by making a monopole-monopole approximation. Value of atomic monopoles are obtained as described above.

4. Static induced dipole energies calculated from an approximate formulation, which becomes more exact as the internuclear distance increases.

Using quantom mechanical and classical mechanical derivations for functions and empirically determined values for some of the derived parameters, atom-atom potential functions are constructed. We then use these potential functions to calculate atom-atom interaction energies as a function of inter-atomic distance.

Finally, the molecular system under consideration is allowed to undergo all possible internal rotations and assume a variety of intermolecular configurations. The energy of each is calculated by summing overall atom-atom interactions. In many instances, the accuracy of these procedures may be checked by comparing calculated results with experimentally determined energies and configurations.

The calculations necessary for these procedures require the development and testing of linked computer programs.

Major Findings:

Approximate intermolecular potential functions can be used quite effectively to predict the energies and conformations of simple molecular systems. They have not yet been tried on large molecular systems (e.g., proteins).

Significance to Biomedical Research and the Program of the Institute:

Accurate prediction of the energies of molecular configurations enables one to know whether specific chemical reactions will or will not occur and why. This information should be of considerable importance to chemotherapy.

Proposed Course of Project:

To develop and extend these methods to more complicated and biologically important molecular systems.

Honors and Awards: None

Publications: None

Serial No. 3.15
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Interview Scheduling for Clinical, Research and Staff Associate Applicants

Previous Serial Number: None

Principal Investigator: Frank A. Petro, Jr.

Other Investigators: Robert J. Connor, Ph.D.; Robert E. Hackman

Cooperating Units: Clinical and Professional Education Section, Clinical Center

Man Years (computed for the 12 month period)
Total: 3.0
Professional: 2.5
Other: 0.5

Project Description:

Objectives:

1. To automate interview scheduling for, and subsequent assignment of clinical, staff and research associate applicants.

2. To develop mathematical algorithms to permit automatic scheduling and assignment by the System 360.

Methods Employed:

1. Matrix algebra and linear programming were used to develop the underlying mathematics.

2. Computer methods required to use the disc and tape of System 360.

Major Findings:

1. The scheduling algorithm has been developed. The mathematical consequences of the algorithm have been demonstrated and are known.

2. Existing linear programming models were found to be adequate for the applicant assignment process.

Proposed Course of Project:

1. The scheduling model has been programmed and is being tested with Clinical Center data. The assignment model is currently being implemented; it will also be tested with Clinical Center data.

2. The automated system is presently working in parallel with the manual system. The old system is expected to be replaced by January, 1968.

Honors and Awards: None

Publications: None

Serial No. 3.16
1. Lab. of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Signal-Noise Relationship in Neuroelectric Data

Previous Serial Number: None

Principal Investigators: Eugene K. Harris, Ph.D., DCRT
Charles D. Woody, M.D., NIMH

Cooperating Units: Laboratory of Neurophysiology, NIMH

Man Years (computed for the 12 month period)
Total: 0.6
Professional: 0.6
Other: 0.0

Project Description:

Objectives:

To characterize nonrandom noise components of neuroelectric signals and, in particular, to distinguish signal-associated from non-signal-associated components of nonrandom noise, with expectation that such discrimination will aid in understanding the statistical organization of adaptive neural systems.

Methods Employed:

A method has been developed (by C. D. Woody) for recovering signal waveforms from background noise by means of an adaptive filter technique using an iterative correlation-averaging procedure. With each iteration, peak crosscorrelations between template and replicate samples determine re-alignment of signals, followed by averaging to produce improved signal in the template record. After each cycle, a computer program provides display of the frequency distribution of deviations of sample waveforms from complete convergence. The average peak crosscorrelation is also computed. Parameters of these frequency distributions, together with observed and theoretical correlation coefficients represent potential tools for distinguishing signal-associated from non-signal-associated nonrandom noise components. Experiments have been performed using both artificial and neuroelectric signals.

Major Findings:

Experiments are not complete but present results indicate that the capability exists for distinguishing and quantifying nonrandom noise components in neuroelectric signals through the proper analysis of relatively simple, computer-displayed results, using the adaptive filter technique.

Significance to Biomedical Research:

These results, if verified by continuing experiments, should provide simple, useful tools for describing dynamic changes in the organization of specific neural systems as monitored by extracellular deep electrodes during processes of sensory stimulation, habituation and conditioning. Further, the adaptive filter technique, useful for recovery of both fixed latency and variable latency signals, may be a powerful technique for summarizing a sequence of repetitive bioelectric signals such as the EKG or EEG, and may thus represent a general pattern recognition device of value in the classification and diagnosis of these important medical signs.

Proposed Course of Project:

(1) Further experimentation with neuroelectric signals in cats under various behavioral regimens; (2) conversion of present LINC computer program to the IBM 360 to achieve much faster operating speeds, to perform statistical analyses within the same program, and to permit use of the Calcomp plotter; (3) experimentation with correlation-averaging technique as a method for distinguishing significant changes in common bioelectric signals under pathological conditions.

Honors and Awards: None

Publications: None (paper by C. D. Woody in press, listed in NINDB report)

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT - 4
Serial Number

2. COMPUTER SYSTEMS LABORATORY

3. A. Demmerle
Acting Chief

I. SUMMARY

This Laboratory, newly activated during this reporting period, is currently pursuing four broad objectives dedicated to the general principles that the computer technology can make substantial contributions in biomedical research in ways other than conventional data processing. These four objectives and the projects through which they are pursued are qualitatively set forth. These projects include the design and development of computer systems to collect and manipulate data and control experiments in real-time, a system dedicated to ease the burden of Clinical Center information processing, systems to collect and process standard types of laboratory data, a system for biological image processing, a hybrid (analog/digital) computation facility, and other generally new applications of computers. All of these programs are in their infancy, but all show considerable promise.

II. CURRENT LABORATORY PROGRAMS

The Computer Systems Laboratory, activated midway through this reporting period, had previously been named the Computer Equipment Laboratory. The new name reflects the overall systems orientation of this group, that is, both hardware and software rather than only a hardware emphasis. The group is presently composed of 25 individuals representing the disciplines of engineering and mathematics. Many of the projects undertaken by these individuals were started late in this reporting period because many of the members of this Laboratory started their employment at NIH during this period.

A. Objectives

This Laboratory is currently dedicated to four broad objectives:

1. To develop and implement project supporting computer systems for the collection, manipulation, and display of biomedical laboratory data

2. To develop and implement an information handling and retrieval system for NIH use which would offer the advantages of the computer and real-time display technology

3. To meet the special data processing systems and equipment requirements of NIH investigators

4. To find and develop new applications for the computer technology in biomedical research

The way in which these objectives have been pursued through projects is summarized below.

B. Progress

The objectives previously set forth represent the current broad goals of this Laboratory. A number of projects have been undertaken in pursuit of each of these objectives. These projects, grouped under their general objectives, are briefly reviewed below. More details for many of the projects are included as Individual Project Reports.

Progress Toward Objective #1 (To develop and implement project supporting computer systems for the collection, manipulation, and display of biomedical laboratory data)

There are many project areas at NIH in which the use of a computer as a tool has not yet been explored or exploited. The laboratory environment is a particularly fertile area in which computer systems can be applied. It is generally our intention to help uncover these project areas, and, by working in cooperation with the scientists involved, design computer systems to meet the requirements of these project areas. These systems, once developed, would be taken over for use (including the responsibility for operation and maintenance) by the project area.

There are at least five general ways in which a computer might be used in biomedical research. They are:

a. Analyzing experimental data which has been collected

b. Simulating processes and making theoretical calculations based on a mathematical model of a system

c. Collecting and analyzing data in real-time by having instruments directly connected to the computer

d. Controlling experiments while they are being conducted, and

e. Developing new ways to do experiments which become possible only by virtue of the computational power of the computer

In an effort to exploit these potentially useful applications, we have looked in detail at the laboratory environment of the NIDR, the Laboratory of Biochemistry of the NCI, and the Laboratory of Metabolism of the NHI. In the first two areas, NIDR and NCI, (described in detail in the Individual Project Reports DCRT 4.1 and 4.2) the uses to which a computer system could be applied were examined in detail, and the optimum computer system to meet these needs was designed. At present, we expect to participate in the procurement and development of the system to meet the needs of NIDR. The NCI computer system, however, seems to be "shelved" due to lack of support.

In the case of the NHI (described in detail in DCRT 4.3), the needs were examined in detail to see how they could be met by use of the CDC 3100 digital computer, a part of the Hybrid Computer System which is presently located in the Clinical Center (Building 10). This system represents a compromise design in order to make use of a computer already owned by NIH, and located in the vicinity of (that is, several floors away from) the Laboratory of Metabolism.

In the final analysis, the useful application of this technology is dependent upon laboratory scientists learning how to use the computer as a tool, since a fundamental appreciation of its capabilities is necessary if it is ever to be used in an innovative way.

Progress Toward Objective #2 (To develop and implement an information handling and retrieval system for NIH use which would offer the advantages of the computer and real-time display technology

The concept of using the computer and reactive display technology in a clinical center environment has been considered for some time, not only at NIH, but elsewhere around the country. Here and there, large investments of manpower and money have been put into the exploitation of this concept. To date, it is not clear to what degree these pursuits are worthwhile, and which areas of hospital and/or research management are best served with these techniques. Our investigations in this field, carried on during this reporting period, lead us to the conclusion that we should proceed with the development of a generalized display system. It would be composed of cathode ray tube type hardware supported by a medium-size digital computer system to manipulate and store data. This system would provide the basis on which any textual or clerical type of information might be displayed and manipulated. With such a general-purpose tool available, ideas can, with relative ease, be implemented and evaluated. One of the first applications to which such a general system might be put is to demonstrate the utility of an automated patient registration system. If a pilot registration system proved worthwhile, a "full-blown" system could be implemented with hardware and software tailored to that purpose, and the utility computer-display system could be applied to the trial of other ideas. Some of these ideas are explored in greater detail in the Individual Project Report DCRT 4.4.

Progress Toward Objective #3 (To meet the special data processing systems and equipment requirements of NIH investigators)

This objective was pursued in three general areas.

The first area was that of offering consultation on the application of computer systems. The largest user of our expertise in this area, during this period, was the NHI Myocardial Infarction Program. Several Myocardial Infarction Centers are to be established around the country to conduct research. This Laboratory provided assistance in the evaluation of proposals and in the contract negotiation for these Myocardial Infarction Centers, particularly as applied to the sizable data management portion of this program.

The second general area is that of establishing the equipment facility necessary to collect certain kinds of data for further computer analysis. We have, during this reporting period, initiated the design and procurement of a portable analog data collection facility and a analog-to-digital data conversion system to prepare data which is collected in an analog form for entry into a digital computer. In the ensuing reporting period, we expect to provide more up-to-date facilities to meet the growing data collection and conversion needs of the NIH scientists who collect relatively small amounts of data during any one experiment, and who have no need for real-time analysis of that data.

The third area is that of computer language development. The list processing language known as LISP which has, in its basic form, been in existence for several years, has not, until this year, been improved and compiled for use on the IBM 360 system. A member of this Laboratory participated in this (primarily IBM) project to make LISP 1.5 suitable for use on the IBM 360 system. This system, and expert advice on its use, will soon be available to NIH IBM System 360 users.

Progress Toward Objective #4 (To find and develop new applications for the computer technology in biomedical research)

This objective overlaps some of the others in scope due to its broadness. The emphasis here is on new applications. Work to meet this broad objective was carried on in four areas: biopolymer sequence analysis, image processing, radiation therapy planning, and hybrid computation.

Work on the problem of computer reconstruction of biopolymer sequences was initiated two years ago and has developed along the lines discussed in Individual Project Reports DCRT 4.5 and 4.6.

An analysis of the NIH requirements for biological image processing was conducted. It made clear the usefulness of a reactive computer-display system for the study and measurement of biological images. It is expected that experimental work with an image processing system would identify specific parameters of possible value for eventual routine processing of such images. On this basis, the elements of the appropriate computer-display system were determined, and serve as the basis of a system which we expect to buy from industry. Once the system is purchased, the software to make it usable to the scientists must be developed. When the system becomes functional, we expect biologists and other scientists with image processing problems to use it collaboratively with the mathematicians and computer specialists of this Division to develop the algorithms and methods applicable to their problems. This project is discussed in greater detail in the Individual Project Report DCRT 4.7.

The Radiation Branch of the NCI has procured a small special purpose computer system to compute isodose curves for therapy planning. This system can be made far more powerful if it is provided with computational support from a larger general purpose computer system. For example, feedback methods can be developed to construct radiation treatment planning without the manual techniques necessarily used now, and, in general, more of the processes of source location and dose calculation can be delegated to a computer system. To do this, however, requires supporting the NCI "Programmed Console" with another

computer of power which is comparable to those in our central facility. This is among the first NIH applications of teleprocessing of data from a terminal other than one with a keyboard; therefore, it is serving as an impetus to the development of a near real-time teleprocessing facility. At the same time, we are applying the mathematical and programming expertise of this Division to help develop this radiation treatment planning system into a useful tool. This project is discussed in a broader context in Individual Project Report DCRT 4.8.

The Hybrid Computer System, a project which was conceived in Fiscal Year 1964, is composed of both an analog and digital computer with suitable analog and digital inputs. It can be a powerful tool for the solution of biomedical problems which involve biological system modeling, and, in general, problems whose solution calls for the advantages of analog computation combined with digital computation. To date, however, it has not met expectations since the system has been plagued with hardware problems which are associated with the analog computer portion of the system. Until these problems have been solved, no progress can be made toward the goals for which the system was procured. Concurrently with our efforts to put the system in operating condition, we have been developing the software utilities essential for the control and use of the system. There is a small queue of potential users from other institutes waiting to use the system in collaboration with this Division. Their problems, such as pulmonary system modeling and coronary system modeling, particularly lend themselves to this type of computational tool. It is possible that the digital portion of this system (a CDC 3100) will be devoted on a part-time basis to other purposes, as has been mentioned under "Progress Toward Objective #1." This decision will be based on the potential utilization of the Hybrid System per se versus other needs to which some of this system's components might be put. This project has been described in Individual Project Report DCRT 4.9.

General Remarks

The majority of the programs described above are not, at present, adequately staffed to assure their proper development. It is our expectation to recruit computer specialists, particularly those with engineering, physics, or mathematics backgrounds, to supplement these efforts. Financial support has been adequate; however, these programs now (and will even more so in the near future) suffer from inadequate laboratory space. It is essential that space be made available for an electronics laboratory, a necessary facility for the development of computer hardware systems, and it is also essential that space be made available to house the image processing facility.

Serial No. DCRT - 4.1
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Dental Institute Computing System

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigators: William Holsinger

Cooperating Units: None

Man Years
Total: .5
Professional: .5
Other: 0

Project Description:

Objectives:
This project is intended to provide on-line real-time computational support for the biochemistry and neurophysiology research programs of the National Institute of Dental Research. The computer system is designed to automate data acquisition from instruments in the biochemistry laboratories, to perform limited computations on captured data, and eventually to control experiments. During neurophysiological research procedures, the system will provide information relative to the proper placement of electrodes, pertinent waveforms enhanced by averaging techniques to improve signal-to-noise ratios, and correlative relationships. The system will provide a test bed for the development of hardware and software techniques for automating many instruments that currently have not been automated on even an off-line basis.

Methods Employed:
The Dental Institute requirements were analyzed in depth, and a system was specified using standard system engineering techniques.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

The proposed system will permit this Division to develop software and hardware for use in similar applications around the campus. Such a system will be used as a research tool in the design of an optimal system for biomedical applications stressing data acquisition. Consistent with the development of time-sharing software on the Central Computer Facility IBM 360/50, the system will provide a basis for analysis and development of satellite computing techniques in a biomedical environment.

Proposed Course of Project:

Upon receipt of program approval, the recommended system configuration will be procured on open bid. Subsequently, instruments will be interfaced with the computer, individual instrument software packages will be generated, and a real-time monitor will be developed. Finally, a communication link with the Central Computer Facility will be established.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.2
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Real-Time Computer System for Biochemistry in NCI

Previous Serial Number: 3.2

Principal Investigator: Marvin Shapiro

Other Investigators: Herbert Sober, Daniel Syed

Cooperating Units: Laboratory of Biochemistry, NCI

Man Years
Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:
The goal is to support research in biochemistry in the National Cancer Institute with a real-time computer.

Methods Employed:
A detailed study of the computer needs of the Laboratory of Biochemistry, NCI, was made. From this, a specific proposal was made for obtaining a small real-time computer system and placing it in that laboratory.

Major Findings:
Implementation of the proposed system is considered to be within the bounds of present technology and of great potential benefit to the Laboratory of Biochemistry.

Significance to Biomedical Research and the Program of DCRT:
The development of computerized automation in a biochemical laboratory could significantly change and simplify present laboratory methods, and lead to the establishment of new analysis techniques.

Proposed Course of Project:
This project has been put aside due to lack of support; however, it can be easily resurrected either for implementation in the NCI or elsewhere. Although the system was designed specifically for the

Laboratory of Biochemistry, NCI, most of the study, and perhaps the entire proposed system could be applied to a similar biochemical laboratory.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.3
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Study of Hybrid Computer Utilization in Support of Laboratory of Metabolism, NHI

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigators: William Holsinger

Cooperating Units: None

Man Years
Total: .25
Professional: .25
Other: 0

Project Description:

Objectives:
This study is designed to investigate the feasibility of expanding the current hybrid configuration to support biochemistry research programs of the Laboratory of Metabolism, National Heart Institute. The ultimate goal of such an expanded system is simultaneous data acquisition and computation capability for both the current mathematical modeling program and proposed biochemistry research programs.

Methods Employed:
Laboratory of Metabolism requirements have been studied, and modifications to the hybrid configuration are being designed using standard system engineering techniques.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:
This study will provide the Division with the necessary data to determine the potential capacity of the Hybrid System to function concurrently as a mathematical modeling station and as a data acquisition

and computational tool for biochemistry applications. Because of readily apparent limitations of the Hybrid System to provide truly concurrent operation as described above, the primary problem under study concerns the determination of the cost, and hence, feasibility of realizing limited concurrent operation.

Proposed Course of Project:

If the cost of modification of the hybrid facility relative to the projected performance of the system in the mathematical modeling and biochemistry areas is considered prohibitive, the project will be terminated, and alternative methods of satisfying Laboratory of Metabolism requirements will be investigated. In the event that expansion of the hybrid facility is deemed feasible, appropriate software will be generated, and additional equipment corresponding to the modification will be procured and installed. A transmission link from the hybrid computer to the Laboratory of Metabolism will also be installed.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.4
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1966

Project Title: Hospital Patient Record and Information Retrieval System

Previous Serial Number: None

Principal Investigator: Victor Colburn

Other Investigators: None

Cooperating Units: None

Man Years
Total: .5
Professional: .5
Other: 0

Project Description:

Objectives:
This effort was initiated in January 1967, and is directed toward application of computer systems and techniques within the environment of the NIH Clinical Center and Research Institutes, for the joint benefit of patient care in the Clinical Center and the underlying biomedical research activities.

Objectives progress in the following manner:

1. The near objective is to study the various facets of the overall "Computers for Health Care" problem, and to determine a logical and practical approach for applying computer technology to specific areas in the Clinical Center in such a manner as to ease work loads and otherwise contribute to improvement of patient care.

2. The mid-range objective (2 to 3 years) is to design and implement an effective and useful system within the Clinical Center which can be extended and interfaced with other elements as better common understanding of the overall problem evolves.

3. The long-range objective (beyond 3 years) is to achieve a broad system network which will effectively handle the greater portion of NIH Clinical Center and Research Activity workload in areas of Data Communication, Record Keeping and Medical Research Information Retrieval.

Methods Employed:
Activity thus far has involved studying efforts of others, gathering ideas, defining problem areas, and determining possible directions for solutions.

Major Findings:
At this point, there are no major findings as such; however, we have arrived about midway in the "near objective" category mentioned above.

The following determinations have been made:

1. The most practical place to start seems to be with the patient registration procedure. The system would be expanded from that point to include an abbreviated patient medical record, bed list, etc.

2. Such efforts should be planned so as to provide for concurrent development of a Medical Research Information Retrieval System.

3. The system (hardware and software) should be highly user-oriented to provide simple operating procedures for non-computer oriented personnel. Software should incorporate multi-level operation, pseudo-conversational modes and user-oriented vocabulary.

4. A functioning demonstration model of such a system should be assembled and evaluated. This might be a one-year program, at the end of which, the subsequent course of the project would be determined.

A preliminary design for such a model has been prepared.

Proposed Course of Project:
Plans have been formulated to proceed with the design and construction of the above described demonstration model during Fiscal Year 1968. If the course of this effort meets our expectations, it seems likely that it will lead to a substantial effort over a 2 to 4 year period starting with Fiscal Year 1969.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.5
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Sequence Determination of Proteins and Nucleic Acids

Previous Serial Number: 3.1

Principal Investigator: Marvin Shapiro

Other Investigators: Marie Chang, Jay Vinton

Cooperating Units: None

Man Years
Total: 1.5
Professional: 1.5
Other: 0

Project Description:

Objectives:
The goal is the mathematical reconstruction of protein and RNA sequences from laboratory data.

Methods Employed:
An algorithm was created for solving the logical problem of deriving from large amounts of laboratory data on an unknown biopolymer sequence as much information as possible about the exact order of subunits in the sequence. The problem is analogous to solving the following letter puzzle: Given the composition of an unknown string of letters (that is, the number of occurrences of each type of letter in the string) and a set of known (short) letter sequences obtained from the unknown string, by overlapping the short sequences, condense the information so that it reduces, if possible, to one string, the one that is unknown. A computer program was written to perform the algorithm.

Major Findings:
The computer program was completed, and over 100 different sets of data (generated by Monte Carlo methods) were run with it. Results show the expected number of fragments of data needed to completely reconstruct sequences of given length and composition. A significant computer result was that beginning fragments of a sequence, which are difficult to identify chemically, need not be identified as such since

the information about what fragment is the beginning one is, in general, derivable from other data.

Significance to Biomedical Research and the Program of DCRT:
Use of the program would enable sequences much longer than heretofore worked on to be tackled. In addition, available data can be handled much more easily with the aid of the computer program than with present hand methods.

Proposed Course of Project:
The computer program will be used with data from alanine transfer RNA and with any other laboratory data made available to us. The program will be modified as required to handle new types of data.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.6
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Computer Applications to the Problem of Determining Molecular Structures

Previous Serial Number: None

Principal Investigator: Marvin Shapiro

Other Investigators: Marie Chang, Jay Vinton, Howard Nash, Micah Krichevsky, Carl Merril, Lewis Sheiner, George Weiss

Cooperating Units: Laboratory of Neurochemistry, NIMH; Laboratory of Microbiology, NIDR; Laboratory of General and Comparative Biochemistry, NIMH

Man Years
Total: 2
Professional: 2
Other: 0

Project Description:

Objectives:
The object is to develop computer programs for solving problems in the determination of the structure of complex molecules, mainly proteins and ribonucleic acids.

Methods Employed:
Mathematical models were developed for characterizing structures in two and three dimensions.

Major Findings:
Some of the applications include the completion of a language (SCOGO) for solving three-dimensional problems in geometry, a program which lines up two given biopolymer sequences in an optimum fashion, and a program to minimize a complex energy function to find a chemical structure with lowest energy. A computer program for fitting a Gaussian type of function to chromatographic data is being written, and will be applicable to all types of chromatographic data where areas under peaks are required.

Significance to Biomedical Research and the Program of DCRT:

The findings of the simulation project could provide a much greater understanding of the work of sequencing and lead to significant reductions in the expenditure of time and equipment. As a result, work on much larger sequences than are now being attacked or even contemplated could be undertaken.

Programs like the one for analyzing chromatographic data will become part of a general library easily accessible to any scientist with digitized data. This will result in great savings of researchers' time, now spent in data analysis, and will also provide more accurate results.

Proposed Course of Project:

A large program has been initiated to simulate the laboratory process of sequencing an RNA. The computer results will indicate the expected sequencing time, will point out bottlenecks to be encountered in the sequencing work for a given RNA, and will indicate ways of optimizing the procedure.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.7
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Biological Image Processing

Previous Serial Number: None

Principal Investigator: Herman Vreenegoor

Other Investigators: Malcolm Bruce, Louis Hodes, Lewis Lipkin

Cooperating Units: Perinatal Research Branch, NINDB

Man Years
Total: 4
Professional: 4
Other: 0

Project Description:

Objectives:

1. Development and utilization of new techniques and devices to permit the study of biological images in order to determine significant characteristics of these images for classification and identification purposes

2. Research directed toward the development of a "natural language" communication system to optimize biologist-computer interactions in the context of image processing.

3. Study, evaluation, and implementation of techniques permitting semi-automatic and/or automatic processing of large volumes of biological images

Methods Employed:

A study was made to learn about existing techniques and devices available for image processing, and to determine what resources would be needed to meet NIH requirements. The study involved site visits, literature searches and meetings with people engaged in the study of biological images.

A unique aspect of the study is the fact that it was undertaken by a multi-disciplinary staff comprised of biologists, engineers, mathematicians, and computer programmers. This composition reflects the

complexity of applying advanced technological concepts to biological problems.

Major Findings:

The study resulted in a recommendation for a reactive computer-display system to aid in the advance of the state-of-the-art of biological image processing. Approval for this system was obtained and resulted in a Request for Proposals for such a system. Proposals are due on June 15, 1967.

The proposed system reflects a determination to concentrate on facilitating the study of biological images with direct, rather than passive, participation of biologists in a multi-disciplinary environment.

Significance to Biomedical Research and the Program of DCRT:

The project will aid biomedical research in that it provides the resources to study biological images in an objective and repeatable manner. Therefore, it should aid the biologist in obtaining a better understanding of the underlying structure. Eventually, it is hoped that techniques can be developed to permit analysis of large quantities of images which would be extremely important for both diagnostic and research purposes. The project is commensurate with the goals of the Division in that it applies advanced computer technology to an important segment of biological investigation.

Proposed Course of Project:

1. a. Evaluate proposals for a computer-display system, and make a recommendation for a specific system.

   b. Prepare the necessary procurement documents, including acceptance criteria.

   c. Determine site requirements and time table to effect possible modifications.

   d. Acceptance testing of the system will commence immediately after delivery.

2. Determine, specify, and implement a sufficiently flexible software display system to facilitate the study of specific biological problems. It is anticipated that this effort will require a contract with a commercial firm for expert advice and for highly skilled (and scarce) programming resources.

3. Implement and evaluate a picture processing oriented software system (called PAX) for the IBM 360/50 computer.

4. Design, implement, and evaluate specific pattern recognition oriented algorithms.

5. Determine the usefulness of a flexible, programmable flying spot scanner as an input device to the system.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.8
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Linear and Non-linear Programming Methods for Biomedical Problems

Previous Serial Number: None

Principal Investigator: William White

Other Investigators: None

Cooperating Units: Radiation Branch, NCI

Man Years
Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:
The objective of this project is to apply linear and non-linear programming methods to some of the problems of the research laboratory, specifically to problems requiring optimization such as the treatment plan in the multi-beam radiation therapy of localized tumors.

Methods Employed:
The methods used are computer techniques using the IBM 360, and with the cooperation of the Radiation Branch, NCI will include a "Programmed Consolé" special-purpose computer. A linear programming code has been written for the IBM 360 as the first part of a program package for the radiation therapy study. Some progress has been made in the coding of a general-purpose quadratic programming code.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:
The application of mathematical optimization has been demonstrated as a powerful tool in business management and engineering applications. Some biomedical problems are concerned with the maximization or minimization of some function. Mathematical programming provides a way to find an optimal feasible solution to this type of problem.

Proposed Course of Project:

It is planned to provide a general-purpose versatile program package to enable the biomedical research worker to try linear and non-linear programming as a solution to some research problems. This work should be completed (except for a small requirement for maintenance) in the next fiscal year.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.9
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Hybrid Computer Project

Previous Serial Number: None

Principal Investigator: Perry Plexico

Other Investigators: Linda Bernstein, Arthur Schultz, David Songco, Robert Romanoff, Philip Turner

Cooperating Units: Section on Clinical Biophysics, Cardiology Branch, NHI

Man Years
Total: 4
Professional: 2
Other: 2

Project Description:

Objectives:

1. To develop a hybrid analog/digital computer system and methods of applying it to the simulation and modeling of biological systems

2. To provide capability for data conversion and editing, and for simple real-time data analysis

Methods Employed:
A hybrid computer system has been purchased and installed. The software and methods necessary for mathematical optimization techniques and for modeling of biological systems are being developed.

Major Findings:
The Hybrid System was accepted by NIH in July, 1966 after which time the analog and linkage portions of the system were expanded to accommodate plans for simulation and optimization of a lung model, as well as work in the general areas of cardiology and neurology. Since its expansion, the analog computer has not been brought to an operating status, due to serious engineering design problems.

A system of computer programs have been developed to accomplish communication between the analog and digital computers and to permit the processing of external interrupts on a real-time basis. In addition, the MIMIC simulation language was adapted for use on the CDC 3100 digital computer. This is a user-oriented compiler language to allow the digital computer solution to problems in a manner similar to that employed on an analog computer.

Significance to Biomedical Research and the Program of DCRT:
This program has not matured sufficiently to have yielded significant useful results.

Proposed Course of Project:
All efforts of this project are being directed toward solving the hardware design problems in the analog computer. When the system is made fully operational, these efforts will be redirected to the development of simulation and modeling techniques.

Honors and Awards: None

Publications: None

July 1, 1966, through June 30, 1967

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT - 5

2. PHYSICAL SCIENCES LABORATORY — 3. Dr. G. H. Weiss

I. SUMMARY

During the present reporting period, the Physical Sciences Laboratory was set up, its objectives were formulated, and some programs were begun. These include a theoretical investigation of the ultracentrifuge and other problems related to chemical separation and measurement, collaboration with other N.I.H. investigators in the planning and analysis of experiments relating to the use of lasers to kill tumors, the design and testing of a new electrocardiograph and the analysis of electrical noise in media with nonlinear properties, as exemplified by nerves.

II. CURRENT BRANCH PROGRAMS

A. Objectives

To develop a group with theoretical interests to collaborate with experimentalists at N.I.H., suggesting experiments based on physical and chemical theories, and to analyze the results of those experiments.

To act as consultants to other members of DCRT who have problems related to the physical sciences.

B. Progress of Current Programs

1. Development of Ultracentrifugation Theory

a. Calculations were completed for the effects of concentration dependent sedimentation on velocity experiments. These calculations suggest most accurate techniques for reducing data from the ultracentrifuge under conditions of nonideality. These are quite important in biological applications.

b. Theory has been developed for equilibrium experiments with continuous slowing. The calculations indicate that substantial savings in time are possible over experiments now performed.

c. Work has begun on techniques for interpretation of ultracentrifugal data from polydisperse systems. Some parallel work is being pursued on similar problems relating to the amino acid analyzer.

2. Study of the Autoanalyzer

a. Several problems relating to the autoanalyzer are being studied by members of this section. The first of these, a study of the relation between sampling intervals for determination of the maximum and minimum of peaks in the output, is nearly complete. It has been shown that it is possible to correct for longer sampling intervals by approximating the peaks by parabolas in the vicinity of the extrema.

b. A study has been initiated into the effects of carryover on peak maxima and mimima. These effects are considerable and lead to substantial loss in accuracy if no account is taken of them. A program for deriving corrections for carryover is presently being written.

3. Consulting Services

a. A collaborative effort between members of NCI and a member of the Physical Sciences Laboratory is underway to determine whether tumor recurrent after laser radiation has growth or other biological properties different from "normal" tumor. These experiments are carried out on mice. Preliminary results indicate that growth rates are equal but that the ability to cause metastases may differ.

b. Scientific planning and statistical consulting services are being provided for a project with Dr. Yashar Hirshaut of NCI on the effects of drugs commonly used in cancer chemotherapy on strains of white cells from acute leukemics grown in vitro. Thus far the experiments appear to provide a reliable technique for measuring time dosage variations on drug effectiveness.

4. Fundamental Studies

a. A project is currently underway to design a new electrocardiograph which makes much more accurate determinations of the heart dipole than present lead systems. The work reported on is being carried out in collaboration with Professor Eugene Fishmann of Howard University who is building a grid electrocardiograph for evaluation of its ability to discriminate between healthy and diseased hearts.

b. Research into theories of various problems in chemical kinetics is being carried out. Present interest is centered on trying to reconcile the deterministic with the stochastic formulation of the theory of reaction rates. Other research is being carried out on models for protein denaturation and renaturation. Further work on this project depends on the availability of experimental data.

c. Studies are being made of models for excitation properties and internal motions of fluids. This is part of a program of investigation of transport properties of materials with the intention of trying to understand physical properties of biological materials. Of particular interest are transport properties of nerves and membranes. Much of this work is carried out in collaboration with members of NINDB.

d. Some work has been done on the theory of absorption spectroscopy in the far ultraviolet.

C. Problems

1. Recruitment.

Work on several projects is proceeding very slowing due to the lack of a full time scientific programmer. It appears to be very difficult to hire experinced programmers in the scientific area, although data processing people seem to be more easily acquired.

D. Program Plans

1. Development of Ultracentrifugation Theory

In FY 1968 a program for the numerical solution of the Lamm equation which is now runing on an IBM 360 at NASA in New York City will be transferred to the 360 at NIH. Work will be continued on the effects of pressure gradients in the ultracentrifuge, band centrifugation, and density gradient centrifugation.

2. Study of the Autoanalyzer

The present study of sampling time on the accuracy of peak determination will be concluded and further work on the carryover problem will be initiated.

3. Consulting Services

Joint projects with members of the NCI will be continued. If results seem to warrant further refinements of studies on time-dose relations on drug effectiveness, we will try to develop theoretical models for possible therapeutic chemotherapy regimes.

4. Fundamental Studies

a. Studies on the grid electrocardiograph will be continued to ascertain the number of grid leads required for an accurate evaluation of the heart dipole. Work on discriminant functions for present lead systems and the grid, will be initiated so that a fair comparison of the two systems can be made.

b. It is expected that data on noise in membranes will be made available for analysis.

c. Projects will be initiated in computer processing of NMR data, the statistical mechanics of lipid layers and allosteric effects in enzyme kinetics by members who are due to join the Physical Sciences Laboratory in the near future.

Serial No. 5.1
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 20, 1967

Project Title: Theory of the Ultracentrifuge

Previous Serial Number: None

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years

Total:	.75
Professional:	.50
Other:	.25

Project Description:

Objectives:

To determine the effects of various factors such as concentration dependent sedimentation pressure, density gradients, variations in rotor speed, and polydispersity on current techniques for determining molecular weights. To devise corrections and new techniques of ultracentrifugation which bypass or eliminate these effects.

Methods:

The methods employed include numerical solutions to partial differential equations and classical analysis.

Major Findings:

It is possible to achieve a considerable speeding up of equilibrium experiments by systematically slowing the rotor

Significance to Biomedical Research:

The ultracentrifuge is a fundamental tool for determining biochemical parameters.

Honors and Awards: None

Publications:

Weiss, George H. (with I. Billick, M. Dishon, M. Schulz, D. Yphantis):The Effects of Rotor Deceleration on Equilibrium Sedimentation Experiments, Proceedings of the National Academy of Sciences 56: 399-404, 1966.

Weiss, George H. (with I. Billick, M. Schulz):On The Calculation of Moments of Molecular Weight Distribution from Sedimentation Equilibrium Data. Journal of Research of the Bureau of Standards, 71A: 13-17, 1967.

Weiss, George H. (with M. Dishon, D. Yphantis): Numerical Solutions to the Lamm Equation III: Velocity Centrifugation. Biopolymers, 1967.

Weiss, George H. (with I. Billick, M. Schulz): Quasi-equilibrium Experiments with Rotor Deceleration. Journal of Physical Chemistry, 1967.

Serial No. 5.2
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Study of the Autoanalyzer

Previous Serial Number: None

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: Ralph J. Nossal, Ph.D.

Cooperating Units: Clinical Chemistry

Man Years

Total:	.75
Professional:	.50
Other:	.25

Project Description:

Objectives:

To provide basic information on the operation of the autoanalyzer so that it can be operated more efficiently and more accurately on a fully automatic basis. To relate the flow processes to observed peak patterns, and to derive corrections for sampling times, carryover, and other factors which contribute to inaccurate operation.

Major Findings:

It is possible to use a sampling time of 10 seconds for peak maxima, with a maximum error in the vicinity of 1.5%. The sampling interval for minima for the same relative error is 5 seconds.

Significance to Biomedical Research:

The autoanalyzer is a widely used instrument in clinical chemistry units of many hospitals in the United States.

Honors and Awards: None

Publications: None

Serial No. 5.3
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Consulting Services

Previous Serial Number: None

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: Ralph Nossal, Ph.D.

Cooperating Units: Surgery Branch, NCI

Man Years
Total: .75
Professional: .50
Other .25

Project Description:

Objectives:
To provide consulting services in biometry, applied mathematics, theoretical physics and chemistry to workers who are primarily in experimental fields.

Methods:
The methods include statistical analysis, theoretical physics and chemistry, and applied mathematics.

Major Findings:
Most of the work done in a consulting capacity this past year was directed towards the investigation of the laser as an oncolytic tool. One problem investigated was that of possible biological differences between tumor recurrent after apparent eradication by laser radiation and "normal" tumor. Careful measurements of growth do not indicate any differences, but more recent experiments on the differential abilities to metastasize do indicate the possibility of a difference. Other work in this area involves the potentiation of laser radiation by other forms of therapy. There are very striking effects produced by the use of x-rays in conjunction with laser radiation on mouse systems.

Significance to Biomedical Research:

The use of lasers as an oncolytic tool is now widely discussed. Fundamental studies of the kind described above are required as preclinical information.

Honors and Awards:

Award for Outstanding Achievement in Mathematics in 1966 made by the Washington Academy of Sciences to George Weiss (jointly with Marvin Zelen).

Publications:

Weiss, George H. (with S. Perry, J. Moxley, M. Zelen): Studies of Leukocyte Kinetics by Liquid Scintillation Counting in Normal Individuals and Patients with Chronic Myelocytic Leukemia, Journal of Clinical Research, 45: 1388-1399, 1966.

Weiss, George H. (with R. Hoye, A. Ketcham):Growth Rate of Experimental Tumor after use of Laser Energy, Journal of the National Cancer Institute, 37: 819-823,1966.

Weiss, George H. (with R. Hoye, J. Gart, A. Ketcham): Potentiation of Laser Oncolysis with Pretreatment X-irradiation, Radiation Research, to appear.

Serial No. 5.4
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1966 through June 30, 1967

Project Title: Fundamental Studies

Previous Serial Number: None

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: Ralph Nossal, Ph.D.

Cooperating Units: Freedmen's Hospital

Man Years:

Total:	.75
Professional	.50
Other:	.25

Project Description:

This project encompasses several lines of investigation, some but not all of which relate to biomedical problems. These include the theoretical study of problems related to the structure of materials, theoretical studies in statistical mechanics, the theory of epidemics, the theory of electrocardiography and the theory of traffic flow.

Major Findings:

A theoretical description of the properties of fluids has been developed and its consequences are being explored. A phenomenological model for the decay of correlations in random walks has been analyzed in an attempt to understand the extent to which approximations currently used in kinetic theory are valid. Studies on the role of carriers in epidemics are being continued, in which the approximation of a large population of susceptibles is made. The design of an electrocardiograph with a large number of leads is being attempted.

Significance to Biomedical Research:

Some of this work, particularly on statistical mechanics, and the theory of traffic flow, may prove to have no immediate biomedical significance. The design of a grid

electrocardiograph might conceivably be a better discriminator of disease than lead systems now in use. Even if this does not prove to be the case, computer studies may lead to a better utilization of present lead systems.

Awards and Honors: None

Publications:

Nossal, Ralph: Momentum Autocorrelation Function for Systems with Finite Spatial Boundaries, Journal of Chemical Physics 45: 1097-1100, 1966.

Nossal, Ralph (with R. Zwanzig): Approximate Eigenfunctions of the Liouville Operator in Classical Many-Body Problems, Physical Review, to appear.

Weiss, George H. (with K. E. Shuler): Exactly Solvable Nonlinear Relaxation Processes. Systems of Coupled Harmonic Oscillators, Journal of Chemical Physics, 45: 1105-1110, 1966.

Weiss, George H. (with K. E. Shuler): The Relaxation of Moments Derived from a Master Equation, Journal of Chemical Physics, 45: 1110-1114, 1966.

Weiss, George H.: The Analysis of Repeated Examinations for the Detection of Occult Diseases, Health Services Research, 2: 272-285, 1967.

Weiss, George H. (with D. Gazis, G. Newell, P. Warren): The Delay Problem for an n-Lane Highway in L. C. Edie, R. Herman, R. Rothery, eds. Vehicular Traffic Science, New York, N.Y., Elsevier, 1967, 267-279.

Weiss, George H.: Pedestrian Queueing at an n-Lane Intersection, ibid, 280-286.

Weiss, George H.:(with J. Gart): Graphically Oriented Tests for Host Variability in Dilution Experiments, Biometrics, to appear.

Weiss, George H.:(with H. Pettigrew): Epidemics with Carriers: The Large Population Approximation, Journal of Applied Probability, to appear.

Weiss, George H.:(with I. Oppenheim, K. Shuler): On the Decay of Initial Correlations in Stochastic Processes, Journal of Chemical Physics, to appear.

Weiss, George H.: Reply to a Query by Buckley, Transportation Research, to appear.

Weiss, George H.: An Introduction to the Statistical Theory of Irreversible Processes, to appear in Irreversible Thermodynamics, H. Henley, ed.

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY
Report of Program Activities
July 1, 1967, through June 30, 1968

ANNUAL REPORT

TABLE OF CONTENTS

Page

Page

July 1, 1967, through June 30, 1968

PUBLIC HEALTH SERVICE-NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT-2 (Serial Number)

2. COMPUTATION AND DATA PROCESSING BRANCH — 3. H. J. Juenemann (Acting Branch Chief)

I. SUMMARY

During Fiscal Year 1968, the conversion effort which, in the previous year, had consumed almost all CDPB resources and efforts was continued, but more significantly it was finally completed. All of NIH's central computing responsibilities have now been converted to the multi-machine IBM 360 configuration. Both Honeywell 800 Computers have been released. In addition, significant advancements were made in modifying the vendor-supplied operating system to provide reliable, efficient and responsive machine servicing to the wide variety of NIH computer needs. A significant start was made toward providing an NIH based system of remote stations accessing to the central computers. Computational and programming support to NIH research and management programs was increased. By virtue of the completion of the conversion effort and by virtue of improved systems performance, the Branch was able to tailor more specific and responsive support to broad categories of NIH computational needs. The mode of advice and assistance to help the NIH user assume greater responsibility for guiding his own computational efforts was successfully emphasized.

Recognizing ever-increasing involvement of computing in NIH's daily life and recognizing maturation of computing into the broader world of NIH's research philosophies, the discontinuation of CDPB as a separate organization focused on revolving fund operations was proposed and (hopefully) approved. This annual report is then the final one for CDPB as an organizational entity spanning the formative years of computing in biomedical research. From Fiscal Year 1954 through 1968, CDPB gave birth to the sometimes faltering computing capabilities of NIH by first placing computers into the service of NIH research and its management. The development of DCRT's capability for fully integrating computing and computational sciences into the every day life of NIH research and its management is both a reason for and, in fact, a fitting epitaph to the organizational entity known as CDPB.

II. BRANCH PROGRAMS

A. Objectives

The Computation and Data Processing Branch provides a central facility for professional programming, computing and data processing services for NIH. It maintains operational liaison with research investigators,

program officials and administrators in the effective use of its computational and data processing services. The Branch performs all phases of centralized computation and data processing, using general purpose digital computers, data conversion equipment, other peripheral equipment and EAM equipment. Develops general purpose, user-oriented computer systems, routines and subroutines; coordinates the use of Branch equipment in both central and peripheral locations; and exchanges technical knowledge and operating experience with other organizations engaged in computation and data processing in support of biomedical research and related activities.

B. Current Programs

During the first quarter of the year, major efforts continued to be absorbed by the dual task of conversion to new equipment and that of modification of vendor supplied software to adopt it to the NIH environment and to convert it into a responsive and reliable system. The majority of the conversion effort was completed at the end of the first quarter of the year and the first of the two Honeywell 800 Computers was released in September, 1967. The few remaining programs were soon converted to the 360 and the second H-800 was released in April, 1968. The contractor who had been involved in the conversion was released. The H-200, which had originally been procured as an input-output adjunct to the two H-800's, continued its role as a media conversion device (paper tape to magnetic tape, reformatting 2nd generation tapes to those required by the 360's, etc.), but in addition, its modest power (and, therefore, modest cost) was increasingly applied to input editing and data preparation tasks formerly performed on EAM equipment.

The second of the Branches' major endeavors for the year, largely successful at year end, has been to mold the vendor-supplied 360 software into the reliable, responsive and expandable system required by the fantastic variety of demands placed on central computer facilities of NIH. The vendor-supplied software system, called Operating System 360, which controls the processing of jobs through the computers, was at the beginning of the year far from adequate to NIH's needs and, in fact, was far below what the supplier had promised. One significant factor which was also typical of a general situation, was that each individual job tied up the whole machine configuration, even when it was only reading input or printing output. These types of limitations, falling short of promises, combined with continued hardware maintenance and delivery problems had lead the Appropriations Committees of both Houses of the Congress to note in their report [illegible] the Fiscal Year 1968 appropriations hearings the critical situation existing at NIH. As a result of our own constant pressure, aided by the Congressional Reports, a major joint effort was mounted between NIH and the vendor. One of the first fruits of this was a significant improvement in hardware reliability and elimination of a large backlog of pending hardware modifications. As a second major result, operating system components were supplied to NIH and modified to suit NIH's systems. These allowed the computers to operate with four simultaneous "streams". While computation of one program is under way, input for the next few jobs

is simultaneously being read in and stored. Results for the last several jobs are also simultaneously being printed, with one output "stream" and one printer devoted to systems messages, diagnostics and small scientific results which have immediate relevance and another output "stream" and printer devoted to those very large printing loads which result from administrative and management data processing. A third significant advancement was that, in spite of the vendor's initial doubts, NIH succeeded in developing the software modifications to enable one disk storage unit to be simultaneously shared and accessed by two different computers. This eliminated the need to store program and procedure libraries redundantly on each machine. Developmental efforts are underway to extend this capability to three or more machines and to include data as well as programs and procedures. A fourth area of intensive software improvement effort was the implementation of geographically remote stations for input and output of computer runs. By year-end two small computers had been attached to the larger computers via telephone lines and it was shown that a wide range of computing jobs could be effectively submitted to the central computers from remote sites. Software developmental efforts are also underway to introduce conversational type computing from remote sites and to use several forms of graphic displays for text output.

C. Program Progress and Accomplishments

1. Analysis and Programming

The Branch function to provide computer programming capabilities in support of NIH research and management functions was able to be accelerated during the year due to the completion of the hardware conversion effort. In the face of ever-growing demands, the effort to transition the programming functions from a role of "service" to one of "advice, assistance and the development of computational tools" was accelerated. The advice and assistance role is a necessary and implicit adjunct to the tool library development role because no tool is useful unless supplemented by technically competent personnel available to inform potential users of its availability, advice on its applicability, assist in its use and most importantly to maintain, modify and expand its capabilities. The following are examples of this new image.

a. Tablemaker II is the Branch's most successful tool development accomplishment. Early in the year Tablemaker II became operational on the new hardware. The system had resulted from a reworking, expansion and improvement of an earlier CDPB-developed Tablemaker I. This tool typically allows a bio-statistician, a program analyst or similar professionals, who have no experience with computers, to take a basic file of data, extract relevant information, recode the control variables or otherwise regroup or re-express the data. The data can be operated on by logical or arithmetic operations to compute frequencies, averages, expectancy tables and then presented along with appropriate descriptive material in tabular form. Heavy use of this tool is made at NIH where a significant portion of the research involves the

deductive analysis of bodies of observed data. While it was a significant accomplishment to provide this tool, it was perhaps more significant that the system can, and is now, heavily and easily used directly by investigators with the central programming staff providing only consultive assistance.

b. During the year a query package was also adopted and expanded from the earlier CDPB developmental effort. The query system provides an easy to use tool for "browsing" and recalling data from any pre-existing data file based on complex logical statements of conditions. One significant aspect of the new version of the query over its predecessors is it can be used with any of a wide variety of file structures including those where the file structure is itself variable.

c. The development of a DCRT Program Library for the 360's came to realization during the year. A Booklet describing over 300 programs available in our program library was prepared and distributed. The major effort has first been concentrated in the areas of statistics and mathematics. Concurrently, work is also going on in the areas of Information Display, Data Processing Procedures and Special Languages. Significantly, all of these programs have been tested, documented and are accompanied by user oriented instructions to make them an attractive and easy to use substitute for reptetive programming each time the same computational situation is encountered.

d. At year end, multiple efforts are underway to build data file creation, editing and updating tools which will operate either in the normal batch processing central facility mode or can be used via remote access points. These file creation and data editing tools will, when viewed in conjunction with the computational library, the query, tablemaker and other standard programs, allow a complete task to be handled using a variety of tools appropriate to each part of the process.

2. Computer Operations

Throughout the fiscal year, the Branch operated two 360/50's and a 360/40. The former were dedicated to the general computational requirements of NIH research and management while the latter was used almost exclusively in support of the administration of NIH's extramural programs. In May a third 360/50 was added as was the capability for two of the 360/50's to share common libraries. The number of productive hours during which these computers are being applied to computational tasks continued to grow. During the 3rd quarter of the fiscal year the average was 425 hours per machine of production time. More indicative of the general improvement are the figures on turnaround time (the elapsed time between the time a programmer submits his job and the time he received results) and secondly, the total number of jobs processed per day. For the daylight shift, when the predominant use of the computer is for testing new programs, the turnaround time was reduced to a 3rd quarter average of 1 and 3/4 hours. During the night shifts, when the primary use of the computer is for large or "production" runs, the average 60 hour turnaround during the first quarter of the year had, by the third quarter, been reduced below 16 so that in almost all cases all work submitted one day is completed by the next morning. At the

same time the number of runs processed daily has increased from 80 to over 450 per day.

3. Computer Training for Scientists

In order to foster the concept of having NIH investigators take added responsibility for programming their own computation relying only on advice, assistance and tools being provided centrally, the training unit of the Branch provided programming training for 155 scientists. The number of scientists doing their own programming rose from last years mid-year figure of 127 to a third quarter figure for this year of 241. During the year the training unit has added to its curriculum instruction in the use of commercially available remote computing services and the Job Control Language of the 360 Operating System. A course in the PL-1 programming language was, at year end, being developed to augment the training in Fortran.

4. Machine Tabulation

During the year the Machine Tabulation function added the Honeywell 200 computer to its Electrical Accounting Machine (EAM), "punched card" and data preparation operations.

a. The Keypunching operations of the Branch continued to operate beyond its capacity throughout Fiscal Year 1968. This, coupled with a shortage of personnel, forced delays in service and necessitated occasional use of outside contracts. Approximately 300,000 cards are punched each month by Branch personnel and approximately 18,000 per month by contractors. It is significant to note that the cost has been consistently lower for inhouse keypunching as compared to contract keypunching. This cost benefit is maintained in spite of the fact that only the more simple high volume jobs can be done on contract.

b. The EAM Unit, located at the Westwood Building, continues to provide specifically tailored punched card service to the Extramural Program Management Offices located in the Westwood Building. This Unit, throughout the year, maintained an outstanding level of service by consistently meeting tight deadlines in spite of continuous personnel shortages. The Westwood operation also includes a do-it-yourself facility for the exclusive use of certain Grants Management Operations. Utilization of this Unit increased constantly during the year as more and more program managers availed themselves of the almost instantaneous assistance available even from relatively unsophisticated equipment. At year end some of the applications currently being processed on Westwood EAM equipment is being evaluated for conversion to the IBM 360 or the H-200.

c. The on-campus EAM Unit at the Central Facility continued during Fiscal Year 1968 to complete their reorientation. Large on-going 360 computer systems which previously also involved EAM processing steps which are gradually being redesignated to reduce the EAM processing operations to those associated with input data preparation and

validation and this form of processing is being redirected to the H-200. Another major segment of the work of the EAM Unit is that which for economic reasons is best processed on small inexpensive equipment. A significant portion of this load has been reoriented to partial or complete H-200 processing by making use of COBOL and a few standard EAM simulator type programs.

5. Field Activities

The Branch continued its support of the NIAID project at the Middle Atlantic Research Unit in Panama. The CDPB staff member there had, in the previous fiscal year, established a punched card (EAM) system which succeeded in gaining control over the large amounts of basic data needed to support the virological research at MARU. During this fiscal year attention was gradually diverted to data analysis and significant use of the computer facilities of a neighboring Federally supported installation.

During the year the Branch provided extensive analysis and programming support to the Clinical Field Studies Unit, NIAMD, located in Phoenix, Arizona.

6. Publications

Gordon, M. H.: DCRT Program Library Booklet, June 1968.
Krichevsky, M. I., Zaveler, S. A. and Bulkeley, J.: Computer-Aided Single or Dual Isotope Channels Ratio Quench Correction in Liquid Scintillation Counting. Analytical Biochemistry, Vol. 22, #3, pp. 442-464, March 1968.

Podolsky, R. J., Shapiro, N. Z. and Zaveler, S. A.: Isotonic Velocity Transients in Models of Muscle Contraction. Federation Proceedings, Vol. 26 #2, Abstract #1659, pp. 553, March-April 1967.

Maloney, C. J., Bryan, S. and Epstein, M.: Computer assisted primary index preparation. J. Chem. Documen. 7, pp. 223-232, November 1967.

7. Administration

As noted in the summary, this fiscal year was the last for the organizational entity of the Computation and Data Processing Branch. The Branch had begun in 1954 as a marriage between some punched card equipment used by the accounting staff and a unit headed by the noted epidemiologist, Dr. Harold Dorn. The Branch continued to rely on EAM equipment until 1958 when the first computer was installed. This was installed almost as an act of faith in the belief that there was a potentially valuable role which computers might play in biomedical research. This potential was slow to develop and that computer, an IBM 650, was used primarily for administrative work. Only a few pioneering investigators dared. Parenthetically, one of these pioneers was to become, in 1966, the Director of the Division of Computer Research and Technology.

The lack of scientific use of the 650 was only partly due to reluctance. It was also due to the nature of that machine. In December of 1960, a Honeywell 800 was installed. It experienced hardware and software troubles, quite similar to the troubles to be experienced in 1966 when the current equipment was installed. However, by late 1961, these troubles had been largely overcome and the scientific use of the H-800 began to increase more rapidly than its administration use. In May of 1963, the growing workload required installation of a second H-800. Recognizing the growth rate for what it was, the start of a new era, a series of studies were undertaken to examine the future needs of NIH. These studies were completed in 1965 and IBM series 360 equipment was selected. A long and sometimes torturous conversion effort was started with the first 360 equipment installed at NIH in April 1966 alongside the two Honeywell H-800's and the H-200 which had been added in 1964. In this fiscal year--in April 1968--the end of the conversion effort was marked by release of the second of the two Honeywell 800's. The conversion from second generation medium scale Honeywell equipment to large scale so-called third generation IBM series 360 equipment thus spanned the period from June 1965 to April 1968. The problems and difficulties with hardware deliveries, hardware performance, software deliveries, software performance, the operating system, the programming contractor, changing and growing NIH workloads during a protracted period of overlap and conversion and many other problems are well known. Now that they are past, they can in retrospect be viewed as having helped to mature the Division.

The history of CDPB's hardware oriented development is paralleled by developments in the analysis and programming components. The current major emphasis on the concept of developing tools which are applicable to a general class of problems has a thread dating back to 1962. Because many NIH scientific computations involve large amounts of data, we enriched the vendor supplied Fortran compiler with a package of about 20 subroutines which provided the scientifically oriented Fortran language with some appropriate data manipulation capability. This package has been continuously used and during this year became operational on the new hardware. In 1964, the Branch developed a query program which although, in its first version, was specific to the NIH Extramural Core Data File, was the progenerator of the improved version completed this year. The Tablemaker system first became operational in 1964 and was soon followed by a generalized file merging tool. The first library of scientific programs was made available in 1964 containing roughly 100 programs, many of which were not fully tested. This compares with more than 300 in the current DCRT Program Library.

Other indicators of the growth of computing at NIH are the fact that in Fiscal Year 1962, the Branch had 103 persons on its staff and a gross value of $1,470,000.00 for the reimbursement of the costs of the services it rendered. In June of 1967, there were 201 persons on duty with the Branch who rendered services valued at $3,800,000.00.

The foundation of DCRT in June of 1966 was prompted in part of share growth in the amount of both scientific computation and management data processing at NIH. It was equally, however, the result of the gradual realization starting perhaps with various committees and studies made since 1963, that the relationship between computing and biomedical research and its management is much more than a relationship of service. Because of this, the accumulated experiences and resources of CDPB were part of the nucleus for the new Labs and Branches of DCRT. Rather than merely saying that the organization known as CDPB ceased to exist in Fiscal Year 1968, it is more accurate to say that CDPB was absorbed into all phases of the larger concept which is DCRT.

July 1, 1967 through June 30, 1968

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Laboratory Activities

1. DCRT-3
Serial Number

2. LABORATORY OF APPLIED STUDIES

3. Eugene K. Harris, Ph.D.
Chief

I. GENERAL SUMMARY

The primary function of this Laboratory lies in the application of mathematical statistics, mathematics and computer technology to areas of laboratory and clinical research at NIH where opportunities exist and the need is clear for productive collaboration between biomedical and mathematical or computer scientists. In the Laboratory's first Annual Report (FY 1967), a number of collaborative projects were identified, particularly in cardiology, neurology, metabolic studies, and clinical chemistry. Planned objectives of almost all of these research projects described in last year's report have been reached. In support of these projects, and for general use at NIH, two large computer programming systems have been completed, one concerned with the management of continuous pressure and flow data obtained during heart surgery and in catheterization laboratories; the other, a comprehensive set of linked programs for time series analysis. Further, the Laboratory has completed independently-generated research in non-parametric statistical methods and has developed specialized computer programs for multivariate analysis, particularly useful at NIH. Considerable progress has also been made on a promising mathematical approach to the determination of molecular structures from fragmentation data after enzymatic digestion.

II. CURRENT LABORATORY PROGRAMS

A. Objectives

1. To initiate or join in sound collaborative research involving the application of mathematical statistical theory and related computer technology to problems of importance in experimental and clinical medicine.

2. To research and test new mathematical statistical methodology or mathematical models useful in the study of physiological and general biological processes of medical significance.

3. To support these research efforts through the development of specialized computer programs to test biomathematical models and facilitate the analysis of laboratory and clinical data, often collected continuously in analog form and requiring analog-digital conversion.

B. Summary of Current Projects

1. Collaborative Studies with Surgery Branch, NHI, and Surgical Neurology, NINDB.

Research studies pursued during this past year have followed upon the perfection of three fundamental tools required for substantive applications of computer science and mathematical methods in these areas. These tools included two large programming systems begun during last Fiscal Year, extended, fully documented and catalogued for routine use on the computer during this Fiscal Year (DCRT Project Nos. 3.5 and 3.9). The first of these was a system of linked computer programs to analyze continuously recorded measurements of ventricular and aortic pressures, aortic flow and the corresponding EKG. This has been supplemented by an extensive array of automated graphic outputs. Recent experimental use of these programs has included (a) measurement of valvular efficiencies by determining energy losses across normal valves and in valves of patients suffering from idiopathic hypertrophic subaortic stenosis; (b) testing and improving current methods for assessing aortic orifice area by detailed examination of dynamic changes affecting orifice area during a single cardiac cycle; (c) calculation of reverse flow in the aorta which may lead to a better measurement of coronary blood flow.

In addition to the cardiac computer programs, these research studies have depended on the use of precise, compact analog recording and editing equipment, specified and installed by this Laboratory during the early part of this Fiscal Year, and now used routinely in both cardiologic and neurologic research studies. This equipment and the computer programs applied to the recorded data, now enable the investigator to undertake on an efficient basis detailed beat-by-beat analyses and comparisons of the dynamic events occurring during heart contraction. For example, in another of the Laboratory's research studies in cardiology (DCRT 3.6), initiated during the previous reporting period, cross-correlation between EKG patterns of consecutive beats in arrhythmias has proven to be a rapid technique for the diagnosis of different kinds of arrhythmias which may provide considerably more discriminating power than current methods based on changes in beat-to-beat intervals alone. Although written in FORTRAN, these correlation programs run in less than real-time (i.e., less than the time required to collect the data on analog tape). Plans are currently under way in cooperation with the Computer Systems Laboratory, DCRT, to incorporate this technique in specially designed, compact hardware for use in the intensive care of cardiac patients.

Studies of electroencephalographic changes in split-brain chimpanzee preparations begun during FY 67 (DCRT 3.7) have been accelerated during the past year through the use of the (portable) recording equipment mentioned above and also the completion, cataloging and documentation of the set of computer programs for time series analysis

(DCRT 3.9) mentioned earlier. In this collaborative study with Surgical Neurology, power spectral analysis was used to study electroencephalographic changes following unilateral section of the brain stem in split-brain chimpanzees. Although gross neurological and social performance did not appear affected, mathematical analysis showed marked changes in distribution of slow and fast frequencies after brain stem section which may reflect differences between the hemispheres in motor dominance. These studies of EEG records in chimpanzees are continuing; in addition, a recent collaborative study undertaken with Surgical Neurology [NINDB NDB(I)-62 SN/OC 913(c)] has involved the collection and analysis of single- and multiple-cell recordings simultaneously from both right and left motor cortexes in animals and in patients with Parkinson's Disease. These studies hope to define more closely the transmission of information following stimulus at lower levels in the central nervous system. Substantial use of the hybrid computing equipment and statistical programs for the analysis of series of events are required.

2. Mathematical Modeling and Statistical Research

During the past year, much more extensive use has been made of the Biological Model Evaluator, a programming system for the development and testing of mathematical models descriptive of physiological processes (DCRT 3.10). Studies in calcium and sodium kinetics have been continued and new projects undertaken this Fiscal Year concerning the transport and utilization of free fatty acids.

One of the important functions of mathematical statistics within this Laboratory in support of computation generally and statistical research in particular, is the development of specialized programs to handle advanced statistical methodology. In recent years, one of the most useful products of this effort has been the random number generator now available for routine use on the IBM 360 system. Building on this technique, a procedure has been developed for generating random samples from normal multivariate populations (DCRT 3.11). In addition, under the guidance of the Statistics and Quantitative Biology Section of the Laboratory, useful programs for discriminant function analysis, including a nonparametric procedure, have been prepared.

One example of an application of mathematical theory (in this case, the algebra of determinants) to solve a management problem at NIH has been the development of an algorithm and corresponding computer program to automate the scheduling of interviews for research, clinical and staff associateships at NIH (DCRT 3.15). Programming and testing were completed in the fall of 1967 and the system fully implemented in January, 1968, for regular operation, during May, 1968, in place of the manual system heretofore used for interview scheduling. It is anticipated that use of the automated system will provide the NIH scientific and clinical staff a valuable 2-3 week additional time period for reviewing applications of associateship candidates.

More fundamental statistical research aimed at increasing the store of advanced statistical methods available to the NIH scientist concerns the analysis of proportions (a continuing and now completed project, DCRT 3.13) and a nonparametric test for the homogeneity of ranges of discrete variables (DCRT 3.14).

In a continuation of previous work on the mathematical determination of molecular structure from fragments obtained after enzymatic digestion (DCRT 3.1) work has been completed on the proof of necessary and sufficient conditions for unequivocal solution of a nucleic acid sequence. A new, supplementary project (DCRT 3.2) in this area concerns an intensive study of graph-theoretic methods, which may lend themselves to the design of relatively simple, efficient algorithms for computer or manual evaluation of fragmentation data.

3. <u>Related Collaborative Studies in Applications of Statistical Methodology</u>

a. Clinical chemistry (DCRT 3.4)

Considerable progress has been made in continuing statistical analysis to determine components of variance in determinations of serum constituents among normal individuals and patients. Analysis of 15 serum variables in 78 normal volunteers is now essentially complete and a series of reports is being prepared on sample selection, laboratory procedures, statistical analyses, and clinical significance of findings, in cooperation with the Clinical Pathology Department of the Clinical Center. When this landmark information is published, it will make available for the first time a well-founded basis for assessing the variability found in multi-phasic screening programs which typically obtain only single samples from each individual. Surprising results have been obtained, for example, regarding the ratios of intra- or inter-individual variation to analytical error, indicating that in some variables, the body exercises a tight physiological control as yet unmatched by the best clinical tests.

b. Signal/"noise" relationships in bioelectric activity (DCRT 3.16)

This project, discussed in the last Annual Report, has succeeded in achieving the goals set for this year. A 360-program for adaptive filtering of signals in band-limited noise has been completed and documented. Supplementing this program are new, efficient programming methods for band-limited noise generation which make use of the random number generator mentioned earlier. These programs have been used to validate a method, conceived last year, for estimating unknown signal/noise levels in prefiltered sample records. Artificial, time-locked sine waves and exponential signals over a 25-fold range of signal/noise level, imbedded in noise of various bandwidths have been subjected to adaptive filtering. Based on statistical analysis of the filtered output, a procedure for characterizing noise patterns in relation to signal content, and for estimating signal/noise ratios has been proven essentially independent of waveform and therefore of wide application

to the analysis of bioelectric signals. For example, it should prove particularly useful for characterizing changes in the organization of background neural activity during the course of habituation to stimuli.

C. Other Activities: Organizational Developments

Laboratory staff members served on the faculty of the Foundation for Advanced Education in the Sciences, teaching two courses: (1) Elementary Probability for the Biological Sciences, and (2) Time Series Analysis.

The most significant change in organization affecting this Laboratory during this reporting period was the recent transfer of the mathematical, mathematical and statistical and library programming units from the Computation and Data Processing Branch to the supervision of this Laboratory.

Note: One research project identified with this Laboratory in last year's Annual Report (DCRT 3.14: Molecular Model Building using Theoretically and Empirically Determined Intra- and Inter-Molecular Potential Functions) has been transferred to the Physical Sciences Laboratory (DCRT 5.4). Another project (DCRT 3.8: A Mathematical Model to Study Cerebral Concussion and Related Brain Trauma) has been incorporated in Project DCRT 3.7: Application of Advanced Mathematical Techniques in Surgical Neurology.

Serial No. 3.1
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Mathematical Problems of Sequence Determination of Nucleic Acids and Proteins

Previous Serial Number: Same

Principal Investigator: James E. Mosimann, Ph.D.

Other Investigator: Jay Vinton

Cooperating Units: Computer Systems Laboratory, DCRT

Man Years (computed for the 12 month period):

Total:	0.2
Professional:	0.2
Other:	0.0

Project Description:

Objectives:

1. To provide mathematical definitions of classes of protein and nucleic acid sequences which cannot be determined using complete digest methods.

2. To provide a probability framework for evaluating the probability that protein and nucleic acid sequences of certain classes are solvable, using partial digest methods.

3. To provide mathematical support necessary for testing computer programs designed to reconstruct nucleic acid and protein sequences from fragment data.

Methods Employed:

The primary mathematical tools are (1) algebra, specifically, the algebra of free monoids; (2) graph theory; and (3) stochastic processes, Markov chains.

Major Findings:

Necessary and sufficient conditions have been obtained for sequences to be solvable using complete digestion fragments. There are broad classes of sequences which cannot be determined by complete digests with monobase specific enzymes. The theoretical studies clearly indicate the need for partial digest fragment data in determining sequences of nucleic acids.

Significance to Biomedical Research:

Determination of the primary structure of proteins and nucleic acids is of extreme importance in studies of the genetic code, as well as molecular bases of life.

Honors and Awards: None

Publications:

Cahiers d' etudes biologiques, Universite Catholique de Lyon, Lyon, France. In press.

Serial No. 3.2
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Graph-Theoretic Analysis of Polymer Sequencing Data Obtained by Fragmentation

Previous Serial Number: None

Principal Investigator: George Hutchinson, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years (computed for the 12 month period):

Total:	0.6
Professional:	0.6
Other:	0.0

Project Description:

Objectives:

To create a mathematical model of experiments in molecular biology which determine protein and nucleic acid sequences by fragmenting the polymer through complete or partial digestion, and determining the fragment sequences.

To develop mathematical criteria by which inconsistencies in experimental data can be detected, by which a polymer sequence or sequences consistent with the data can be computed, and by which all alternative possibilities can be eliminated.

Methods Employed:

Fragment sequences from an unknown protein or nucleic acid sequence are presumed given. A mathematical model using a modification of the theory of graphs (i.e., networks) was developed. A special graph is constructed from any given data configuration of a certain class. Inconsistencies of the data are revealed during the construction process or by simple tests on the constructed graph.

The polymer sequences consistent with the given data are in one-one correspondence with the Euler circuits of the constructed graph. If an Euler circuit is known, the corresponding polymer sequence can be recovered immediately. Therefore, the constructed graph holds the complete solution of the problem in a compact form.

A method of generating all of the Euler circuits of any such graph has been developed. Euler circuits are efficiently generated one by one, and the process can be cut off if the number becomes too large. So, it is possible to efficiently recover all polymer sequences yielding the given data, unless there are too many.

Major Findings:

The graph-theoretic model has been developed for the common situation in which complete digests by pancreatic ribonuclease and by T_1 ribonuclease are available. Complete analysis of data of this type can be made. Criteria have been developed to determine whether the data are inconsistent, to determine the polymer sequence of sequences consistent with the data, and to eliminate all alternative possibilities. Efficient algorithms are available for these computations. The present results are primarily useful in detection of inconsistencies, since two complete digests do not usually provide enough data to uniquely determine a polymer sequence.

Significance to Biomedical Research and the Program of the Institute:

Successful development of this approach would aid researchers in molecular biology to quickly and accurately evaluate experimental data, to detect erroneous data, and to avoid redundant laboratory effort.

Proposed Course of Study:

To extend the graph-theoretic model to accommodate more general configurations of complete and partial digest data.

To design and prepare a computer program or programs for evaluation of polymer fragment data by the graph-theoretic method.

Honors and Awards: None

Publications: None

Serial No. 3.4
1. Lab. of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Statistical Research in Clinical Pathology

Previous Serial Number: Same

Principal Investigator: Eugene K. Harris, Ph.D.

Co-Investigators: Ernest Cotlove, M.D., Clinical Pathology Dept., CC
Donald Young, M.D., Clinical Pathology Dept., CC
George Shakarji, LAS, DCRT

Man Years (computed for the 12 month period):

Total:	.85
Professional:	.75
Other:	.10

Project Description:

Objectives:

To explore, through careful application of statistical methodology, relationships inherent in large volumes of results of chemical analyses routinely performed on serial blood samples from normal volunteers and patients at the Clinical Center.

Methods Employed:

At this time, blood chemistries from 78 normal volunteers, in ten groups of seven or eight individuals per group, are stored in easily retrievable form on digital magnetic tape. These data consist of concentrations of at least 15 elements and compounds (including some enzymes) measured in each of 10-12 weekly blood samples from each subject. The total ten-group data base covers a 2 1/2 year period of study, and includes approximately 26,000 measurements.

Similarly stored patient blood data are available from approximately 2500 Clinical Center admissions during a 9-month period in 1964-65. Demographic and diagnostic information on these patients is also available for related study. Serial blood data for each patient is less abundant than for normal volunteers and not regularly spaced, but does permit some time-wise analyses relative to effects of therapy or other changes of state.

A separate component of variance analysis has been applied by computer program to the weekly concentrations for each normal volunteer. The availability of daily measurements on pooled frozen serum allowed separate estimation of analytical error, so that weekly variation in a subject's results can be apportioned to (a) inherent physiological variability, (b) fluctuations in routine laboratory analysis of replicate samples. Analysis of variance of the subject means permitted estimation of inter-individual variance components.

Major Findings:

During the past year, variance component analysis in the normal control data has been completed, as have many auxiliary analyses testing age-race-sex differences, as well as factors related to laboratory quality control, and to the validity of various statistical estimates. A series of reports concerning sample selection, laboratory procedures, results of statistical analysis and their significance for clinical medicine are now in the final stages of preparation.

Significance to Biomedical Research:

These studies will (1) add substantially to existing information on physiological variation of blood constituents within a single healthy individual over time and among normal individuals within a given age, sex, and race group, and (2) provide new knowledge on currently achievable analytical precision (e.g., in the course of multiphasic screening programs), and (3) provide an opportunity to test potentially useful mathematical models of time-dependent changes in blood constituents of patients during the course of therapy.

Proposed Course of Project:

Further analysis of data from normal subjects will study possible inter-relationships among different blood constituents. One specific aim will be the development of a function of several variables which may provide a distinctive blood "profile" of a normal individual. Later, attention will be centered on the patient file, using distributions of blood tests in normal volunteers to aid in isolating extreme ranges in each variable among patients. Correlations among extremes and relationships to diagnosis and to demographic variables will be examined, and relevant physiological hypotheses will be tested. Finally, the variation in each test over time will be studied in relation to early and final diagnoses and course of therapy. At this point, more sophisticated mathematical statistical methods, drawn from the theory of stochastic processes, will be needed to describe dynamic changes in test results during hospitalization.

Honors and Awards: None

Publications: None

Serial No. 3.5
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Heart Surgical Suite Project

Previous Serial Number: Same

Principal Investigator: David B. Gilbert, M.D.

Other Investigators: Thomas F. Lantry, LAS, DCRT
Martha R. Horton, LAS, DCRT

Cooperating Units: Dr. Andrew G. Morrow, Surgery Branch, NHI
Dr. Stanton Nolan, Surgery Branch, NHI

Man Years (computed for the 12 month period):

Total:	1.7
Professional:	1.7
Other:	0.0

Project Description:

Objectives:

The goals of the Heart Surgical Suite Project are to provide accurate, well-organized mathematical analyses of massive quantities of physiological data obtained from humans during surgery.

Methods Employed:

Data acquired in analog form on magnetic tape during surgery are edited off-line on equipment designed by the Division. The data are then converted to digital form on the CDC 3100. The digital tape is analyzed off-line on the IBM 360 system utilizing FORTRAN and assembly language programs. The parameters presently include two pressures, the flow between these two pressures, and the electrocardiogram. Classical analyses of this data include the maxima and minima of flow and pressure, rates of flow and pressure change, stroke volume, stroke work and power. Experimental analysis includes the energy loss across a diseased valve expressed as valvular efficiency and the distribution of the fractional energy loss throughout a cycle. Reverse flow in the aorta is also calculated.

Major Findings:

The FORTRAN programs for off-line analysis of pressure-flow characteristics were completed and checked out. The data acquisition and display hardware was completed and operational; they were used almost daily not only for the DCRT-NHI project but for DCRT-NINDB projects as well.

Utilizing the hardware and software developed, the non-steady-state characteristics of post premature contractions in idiopathic hypertrophic subaortic stenosis has been documented and quantitated. The nonlinear relationship between empirical indices (gradients valve areas, tension-time indices) and the actual disturbed flow energy losses have been described.

Using selected portions of the developed software, the relative efficiences of mitral ball and disk valves have been studies in the calf. Optimum cardiac outputs for minimal frictional energy losses were defined.

Significance to Biomedical Research:

The development of a hardware and software package for the beat-to-beat analysis of pressure-flow relationships on a digital computer system has been completed. The mathematical, engineering and programming have been documented in detail and is available for other investigators.

Proposed Course of Project:

The completed software and hardware package is now available for NHI-DCRT collaborative projects. The programs are most suitable for describing non-steady-state phenomena in both the animal laboratory and the surgical suite environment. Various models of obstructed flow will be tested and analyzed with the programs; they will be compared to the results observed in vivo. A great bulk of the programs await implementation in the cardiac catheterization laboratory.

Honors and Awards: None

Publications: None

Serial No. 3.6
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Computer Modeling of Arrhythmias

Previous Serial Number: Same

Principal Investigators: David B. Gilbert, M.D.
Eugene K. Harris, Ph.D.

Other Investigators: None

Cooperating Units: Surgery Branch, NHI (to supply data in the form of magnetic tapes)

Man Years (computed for the 12 month period):

Total:	0.0
Professional:	0.0
Other:	0.0

Project Description:

Objectives:

The immediate goal of this project is to simulate the sequential changes expected in cardiac conduction during rapid stimulation using the digital computer. Ultimately, the simulation of sequential electrical activity should aid in pattern recognition studies for classification of arrhythmias.

Methods Employed:

Utilizing known facts about transmembrane potential and thresholds of response for each of the pathways of cardiac conduction, a digital computer simulation of cardiac conduction has been programmed. The response of each nerve fiber is determined by its previous response and the input impulse. For a given number of cardiac cycles, the model will stimulate and plot the action potential of the sinoatrial node pacemaker, and atrial fiber conducting an impulse to the atrioventricular node and the effect of the A-V node upon a ventricular fiber.

Major Findings:

Due to pressure of other work, this project was suspended during FY 68, but will be resumed during the coming year.

Significance to Biomedical Research:

A general need is recognized for the adequate training of physicians to interpret cardiac arrhythmias. The proposed model graphically depicts each step of cardiac conduction as well as the expected time at which one would expect the atria or ventricle to fire. More specifically, the model should aid to our knowledge about the ventricular response to atrial fibrillation and how that response might be altered.

Proposed Course of Project:

The model is to be tested systematically for its ability to predict known changes recognized on the EKG. Data from individuals in known atrial fibrillation will be collected, and the resultant frequency distributions of the R-R interval will be compared to the digital computer simulation. The effects of respiration upon that distribution can be then studied from both a theoretical and experimental point of view.

Honors and Awards: None

Publications: None

Serial No. 3.7
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Application of Advanced Mathematical Techniques in Surgical Neurology

Previous Serial Number: Same

Principal Investigator: Raymond Mejia

Co-Investigator: Dr. Maitland Baldwin

Other Investigator: Thomas F. Lantry, LAS, DCRT

Cooperating Units: Surgical Neurology Branch, NINDB
Technical Development Section, NINDB

Man Years (computed for the 12 month period):

Total: 1.0
Professional: 0.5
Other (including computer time): 0.5

Project Description:

Objectives:

The objective is to study physiological parameters available during surgery through the application of advanced digital processing techniques. Experiments using lower primates have been conducted, and studies are being carried out which include the mathematical analysis of patient data obtained in the Neurology recording facility.

Methods Employed:

LINC computer programs have been developed to accept analog data consisting of up to six physiologic variables and convert them to digital data in a form directly acceptable to the IBM 360 computer. Programs for the IBM 360 computer have been developed which quantify slowly changing variables (such as temperatures), continuously changing phenomena (such as EEG) and discrete variables (pulse rate, for example).

Major Findings:

In collaboration with Dr. Maitland Baldwin, Surgical Neurology, data was obtained and analyzed to study the effects of unilateral section of the brain stem after mesial cerebral incision[1/] [NDB(I)-54 SN/OC 101(c)]. Chimpanzees subjected to mesial cerebral incision are capable of synergistic and other usual motor performance as well as normal social reactions despite severence of interhemispheral connections. Moreover, subsequent subtotal hemisection of the brain stem does not materially alter neurological or social performance nor is it followed by disturbance of nocturnal-diurnal cycles. However, both operations were followed by marked electroencephalographic changes. After mesial cerebral incision, there is an evident asymmetry while following brain stem section, the record is comparatively slower particularly on the side of section.

Significance to Biomedical Research and the Program of the Division:

In addition to the contribution to neurology and the art of surgery, the quantification of meaningful physiological parameters in a manner amenable to immediate as well as long term scientific study is significant. The development and application of suitable mathematical techniques of analysis is made possible in large part through the availability of adequate recording and computing equipment.

Proposed Course of Project:

A study of electroencephalographic changes due to subconcussive and concussive blows, which are measured on the cortex and in deep structures of the brain, is being conducted in collaboration with Drs. A. Ommaya and D. Gainsburg [NDB(I)-62 SN/OC 907(c)]. A study of the relationship between hemispheral and motor dominance is being conducted in collaboration with Dr. Baldwin.

Honors and Awards: None

Publications: None

1/ Reported to the New York Academy of Sciences meeting on the Use of Primates in Experimental Surgery, September, 1967.

Serial No. 3.9
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Time Series Analysis Package

Previous Serial Number: Same

Principal Investigator: Raymond Mejia

Other Investigators: None

Cooperating Units: Computation and Data Processing Branch, DCRT

Man Years (computed for the 12 month period):

Total: 1.8
Professional: 1.3
Other (including computer time): 0.5

Project Description:

Objectives:

There has been a need at NIH for a library of computer programs to aid the analysis of time-dependent data. Benefits from such a set of programs will assure to the researcher who has collected small amounts of data over a relatively long period of time as well as to the project in which real-time data analysis is required to control physical or chemical processes.

A wide class of time series techniques has been implemented using the IBM 360/50 computer at NIH. These programs will be made available to remote users when such terminals are supported by the system.

Methods Employed:

A system of programs for the mathematical analysis of time series has been written, adapted for the NIH computing system and implemented under the current batch processing monitor.

Programs for "raw" data processing include the following:

a. acquisition of single channel and multiplexed data

b. probability density and joint probability density
c. decimation
d. normalization to zero mean and one standard deviation
e. test for linear trend and detrending
f. digital filtering and prewhitening.

Data generators for periodic, aperiodic and random data are available. Programs for correlation analysis include the following:

a. auto- and cross-covariance
b. auto- and cross-correlation
c. higher order correlations
d. lag windows including Parzen's, Bartlett's and Bram's.

Programs for frequency analysis include the following:

a. fast Fourier transform (SHARE routine PK FORT)
b. harmonic analysis
c. spectral and cross-spectral analysis
d. spectral windows including Hanning's and Daniell's
e. coherency between channels.

In addition to card input, data formats are accepted from the LINC computer operated by the Technical Development Section, NINDB, the DCRT 3100 hybrid computer and the SDS 910 digitizer at the Naval Ship Research and Development Center.

In addition to printed output, plots of desired quantities are provided using the DCRT Calcomp plotter.

Documentation in the form of a user's manual is currently in preparation and should be completed by the end of the Fiscal Year.

Significance to the Program of the Division:

The time series package serves as a tool in the analysis of data obtained during investigations undertaken within the Laboratory of Applied Studies and the other laboratories of the Division and the Institutes. It will facilitate the training of biomedical scientists in mathematical techniques which are applicable to their research, and in techniques which will facilitate the utilization of large-scale computation in their research programs.

Proposed Course of Project:

It is planned to continue to develop these programs in two directions. First, to accept problems which are currently too large to be processed by allowing for more input data (with more flexible input formats) in the available computer memory storage. Second, to serve a wider class of problems by adapting to the DCRT computing system when it permits remote utilization.

The optimization of calculations which are particularly time-consuming will be undertaken. The elimination of input parameters which are difficult to compute by the user is also desirable.

After testing and supervised utilization for some time, the package will be made available to other biomedical research activities.

Honors and Awards: None

Publications: None

Serial No. 3.10
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Biological Model Evaluator Studies

Previous Serial Number: Same

Principal Investigator: John E. Fletcher

Co-Investigator: Richard Shrager

Cooperating Units: None

Man Years (computed for the 12 month period):

Total:	1.35
Professional:	1.25
Other:	.10

Project Description:

Objectives:

The object of the Biological Model Evaluator Studies is twofold. The primary object is to collect and develop those computer-oriented mathematical and statistical tools most applicable to the testing of biomathematic models. These tools are to be assimilated into user oriented program packages which provide NIH with a generalized biological model evaluator capability. Secondly, this project acts collaboratively in developing and supplementing mathematical-biological model building in other Institutes at NIH. In so doing, the requirements of biomedically oriented investigators for model evaluation are kept current and the capability of the evaluator programs can be updated accordingly.

Methods Employed:

1. The central concept of this project is model fitting by means of mathematical techniques. Two systems are currently operational. The first is the SAAM (Simulation Analysis and Modeling) system developed by the mathematical research branch of NIAMD, and the second is a nonlinear curve fitting program developed by Mr. Richard Shrager, Physical Sciences Laboratory, DCRT. The SAAM system is essentially a closed-package system of computer programs built

around the concept of compartmental analysis and its associated techniques. This system cannot be used in connection with other programs, and special modifications are necessary for all but a small class of well-defined mathematical models.

The second system is a class of generalized subroutines (NIHH22-23) designed to do nonlinear least squares curve fitting for a wide class of mathematical models. To use this system, a simple FORTRAN program is written defining the proposed model. The programming system accepts any model expressible in FORTRAN notation. In addition, it may be used as a subsystem in a larger user-oriented program. These newer, more versatile techniques are now documented and are being distributed for general NIH use. Further development of this project will permit the inclusion of data handling subroutines or numerical analysis subroutines which may enrich its utility as a biomedical tool.

2. Collaborative efforts are directed to the following projects:

NIH - NCI: Metabolism Branch; Dr. James Phang. The Quantitative Determination of Calcium Kinetics and Bone Synthesis in Normal Subjects, (Ending July 1, 1968).

NIH - NIMH - LCS: Unit on Psychosomatics; Dr. Leslie Baer. The Quantitative Determination of Sodium Kinetics in Normal and Diseased Patients. (Ended Jan. 1., 1968).

NIH - NHI: Section on Polypeptide Hormones, Laboratory of Molecular Disease; Dr. John Potts. Parathyroid Response to Changes in Serum Ca Level.

NIH - NHI: Section on Lipid Metabolism, Laboratory of Metabolism; Dr. Arthur Spector. Free Fatty Acid Transport and Utilization.

NIH - NCI: Laboratory of Biology; Dr. Fred Mushinski. Genetic and Biochemical Mechanisms Controlling Protein Synthesis in Well-Differentiated Neoplastic Cells, Mouse Plasma-Cell Tumors.

NIH - NIAMD: Laboratory of Physical Biology; Dr. Nobuko Kataoka and Dr. Hideo Kon. The Electron Spin Resonance of Low Spin Isocyanide Complexes of Co(II). I. Halides.

NIH - NIAMD - NHI: Laboratory of Physical Biology (NIAMD) and Laboratory of Biochemistry (NHI); Dr. Elliot Charney (NIAMD) and Dr. Lin Tsai (NHI). Spectral and Optical Activity of Conjugated Diketones.

NIH - NCI: Laboratory of Biochemistry; Dr. Herbert A. Sober. Optical Rotatory Dispersion Properties of Oligolysines.

These projects are discussed further in the reports of the respective Institutes. The participation in these efforts has involved basic formulation of an appropriate model for clinically acquired experimental data, data processing of the data and model, and interpretation of results. Refinements of the data collection, experimental methods, and biological models are often suggested as a consequence of mathematical results obtained through the computer program.

Proposed Course of Project:

Continuing efforts will (1) develop a library of special purpose models; (2) catalog supplementary subroutines which increase the utility of the model evaluator system; (3) document educational materials for training prospective users in mathematical-computer tools; and (4) provide collaborative and consulting efforts to selected NIH projects.

Honors and Awards: None

Publications:

Spector, Arthur A., and Fletcher, John E.: The binding of long-chain fatty acids to bovine serum albumin. J. Lipid Research. In press.

Shrager, Richard I., and Fletcher, John E. A User's Guide to Least Squares Model Building. DCRT Technical Report #1.

Serial No. 3.11
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Computer Simulation Methods and Related Statistical Methodology

Previous Serial Number: Same

Principal Investigators: Robert J. Connor, Dr. Eng.
Richard Greenstreet

Other Investigator: David VanSant

Cooperating Units: CDPB, DCRT

Man Years (computed for the 12 month period):

Total:	.96
Professional:	.85
Other:	.10

Project Description:

Objectives:

To provide NIH with special computer-related statistical models.

Major Findings:

1. A procedure for generating random samples from normal multivariate populations with specified mean vector and covariance matrix has been developed. It uses the random number generator reported in last year's Project Report Serial No. 3.11. It is programmed and available on system 360.

2. A discriminant function package has been developed to determine linear and quadratic discriminant functions for the two population case. Also, a non-parametric procedure for discrimination, based on order statistics, has been included (see Kendall[1/]).

[1/] Multivariate Analysis. Proc. of an International Symposium held in Dayton, Ohio, June 14-19, 1965. Edited by Paruchuri R. Krishnaiah, Academic Press, New York, 1966, pp. 165-185.

Currently, work is under way to investigate Kendall's procedure and some modifications of it.

A test for equality of covariance matrices has been programmed in BASIC for the time sharing units. It will be programmed for the 360 system in the near future and fully documented. The properties of this test are being investigated and much of the mathematical analysis essential for this investigation has been performed.

Honors and Awards: None

Publications: None

Serial No. 3.12
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: EKG Modeling and Analysis

Previous Serial Number: Same

Principal Investigators: David B. Gilbert, M.D.
Robert E. Hackman, LAS, DCRT

Other Investigators: None

Cooperating Units: Surgical Branch, NHI (data alone)

Man Years (computed for the 12 month period):

Total:	0.5
Professional:	0.5
Other:	0.0

Project Description:

Objectives:

This year's efforts have been diverted to the technical problems of EKG analysis, specifically toward reducing the high data rates associated with the patient monitoring of electrocardiographic arrhythmias.

Methods Employed:

An algorithm was developed to characterize the uniqueness of a single QRST waveform when compared to a previous or standard complex. Combinations of the R-R interval and similarity of the waveform shape were compared with premature ventricular contractions, A-V block, nodal rhythms and sinus arrhythmias.

Major Findings:

Correlation coefficients in addition to R-R intervals appear to be a promising technique in reducing the high data rates associated with computerized monitoring of the EKG. Arrhythmias may be edited according to general mechanisms of pacing abnormalities and/or conduction abnormalities. Similar steady state normal or

abnormal beats are deleted from the edited string. The algorithm as operated on the IBM 360/50 system operates better than real-time in FORTRAN IV.

Significance to Biomedical Research:

Present evidence points toward single heart beats as containing significant information in regard to acute myocardial infarction. Predictive algorithms of impending electrical failure must necessarily begin with optimal (every beat) analysis before decisions may be made as to the reliability of intermittent sampling. Economic computer every-beat analysis necessarily awaits an effective low-cost preprocessing system.

Proposed Course of Project:

A feasibility study will be made to determine whether the algorithm may be implemented in an 8K 1.5μ second cycle time desk-top machine. Both presently successful and future experimental algorithms will be simulated in 12- and 16-bit word form on the IBM 360/50. The overview will be to determine whether low-cost preprocessing of the EKG may be presently purchasable or fabricated within the Division.

Honors and Awards: None

Publications:

Gilbert, D.B., and Hackman, R.E.: A Data Reduction Technique for On-Line Arrhythmia Analysis. Abstract, Proceedings of the 3rd Joint Meeting, CSCOA, March, 1968.

Serial No. 3.13
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: The Statistical Analysis of Percentage or Proportional Data

Previous Serial Number: Same

Principal Investigator: James E. Mosimann, Ph.D.

Other Investigators: Robert J. Connor, Dr. Eng.

Cooperating Units: None

Man Years (computed for the 12 month period):

Total:	.87
Professional:	.67
Other:	.20

Project Description:

Objectives and Methods:

1. To provide mathematical and statistical models to permit the analysis of data which are proportioned, e.g., per cent serum protein constituents of blood.

2. To provide mathematical support for biological studies of growth and differential growth. The primary mathematical methods involved are those of probability and mathematical statistics along the lines of the multivariate beta or Dirichlet model (Mosimann, 1962, 1963).

Major Findings:

Definitions of independence for proportions are developed. A general class of multivariate distribution whose variables meet the definitions has been developed. The general class describes a competitive situation in which one particular proportion has a dominant role to some other proportions. This means, for example, that one variable assumes its share of a resource without reference to the amount the second variable then subsequently assumes. The models include as a special case the Dirichlet model. Hence, they provide alternative hypotheses to the simpler non-competitive hypothesis associated with the Dirichlet model. In analyzing data

on differential growth published by Mosimann, the models have proved to be of biological use in identifying which of several variables shows gr wth dominant to other variables. Data on bone composition of rats has been analyzed. These data are included in a paper "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution" which has been completed and submitted for publication.

References:

1962. On the compound multinomial distribution, the multivariate beta distribution and correlations among proportions. Biometrika. Vol. 49, No. 1-2, p. 65-82.

1963. On the compound negative multinomial distribution and correlations among inversely sampled pollen counts. Biometrika. p. 47-54.

Honors and Awards: None

Publications: None

Serial No. 3.14
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: A Range Distribution and Related Tests for Homogeneity

Previous Serial Number: None

Principal Investigator: Robert J. Connor, Dr. Eng.

Cooperating Units: None

Man Years (computed for the 12 month period):

Total: .17
Professional: .12
Other: .05

Project Description:

Objectives:

To provide NIH scientists with statistical methodology for testing the homogeneity of events on a discrete scale (or over classes).

Major Findings:

The distribution of the sample range from a finite discrete rectangular distribution has been derived. Utilizing this result a non-parametric test for homogeneity over a discrete scale has been developed and studied. It has been found that this test when used in conjunction with the Kolmogorov-Smirnov test is of considerable value, being sensitive to situations at boundaries of classes.

Honors and Awards: None

Publications: None

Serial No. 3.15
1. Lab. of Applied Studies
2. Stat. & Quan. Biol. Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Interview Scheduling for Clinical, Research and Staff Associate Applicants

Previous Serial Number: Same

Principal Investigator: Frank A. Petro, Jr.

Other Investigators: James E. Mosimann, Ph.D.

Cooperating Units: Clinical and Professional Education Section, Clinical Center

Man Years (computed for the 12 month period):

Total:	2.00
Professional:	1.75
Other:	0.25

Project Description:

Objectives:

1. To automate interview scheduling for, and subsequent assignment of, clinical, staff and research associate applicants.

2. To develop mathematical algorithms to permit automatic scheduling and assignment by the System 360.

Methods Employed:

1. Matrix algebra and linear programming were used to develop the underlying mathematics.

2. Computer methods required to use the disc and tape of System 360.

Major Finding:

The scheduling algorithm has been developed. The mathematical consequences of the algorithm have been demonstrated and are known.

Proposed Course of Project:

The scheduling system is fully implemented. It replaced the manual system in January, 1968.

Honors and Awards: None

Publications: None

Serial No. 3.16
1. Lab. of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Signal-Noise Relationships in Neuroelectric Data

Previous Serial Number: Same

Principal Investigators: Eugene K. Harris, Ph.D., DCRT
Charles D. Woody, M.D., NIMH

Cooperating Units: Laboratory of Neurophysiology, NIMH

Man Years (computed for the 12 month period):

Total: 0.6
Professional: 0.6
Other: 0.0

Project Description:

Objectives:

To characterize band-limited noise components of neuroelectric signals and, in particular, to distinguish signal-associated from signal-unassociated components of such noise, with expectation that this discrimination will aid in understanding the statistical organization of adaptive neural systems.

Methods Employed:

A method has been developed (by C.D. Woody) for recovering signal waveforms from background noise by means of an adaptive filter technique using an iterative correlation-averaging procedure. With each iteration, peak crosscorrelations between template and replicate samples determine re-alignment of signals, followed by averaging to produce improved signal in the template record. After each cycle, a computer program provides display of the frequency distribution of deviations of sample waveforms from complete convergence. The average peak crosscorrelation is also computed. Parameters of these frequency distributions, together with observed correlation coefficients represent potential tools for distinguishing signal-associated from signal-unassociated band-limited noise components. Experiments have been performed using both artificial and neuroelectric signals.

The former have been time-locked signals of two different waveforms in noise of varying bandwidths, over a wide range of signal/noise ratios. Noise generation, signal insertion and adaptive filtering have all been performed digitally on the IBM 360/50 computer.

Major Findings:

Results obtained during the present reporting period show conclusively that the capability exists, essentially independent of signal waveform, to quantify signal-associated noise components in neuroelectric signals, and to estimate pre-filtered signal/noise levels, through statistical analysis of computer output from adaptive filtering.

Significance to Biomedical Research:

These results provide simple, useful tools for describing dynamic changes in the organization of specific neural systems as monitored by extracellular deep electrodes during processes of sensory stimulation, habituation and conditioning.

Proposed Course of Project:

(1) Further experimentation with neuroelectric signals in cats under various behavioral regimens; (2) computer studies of artificial, variable-latency signals to determine the feasibility of estimating the distribution of latencies by application of mathematical models to output from adaptive filtering.

Honors and Awards: None

Publications:

Woody, Charles D.: Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. and Biol. Eng. 5: 539-553, 1967.

July 1, 1967 through June 30, 1968

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT - 4
Serial Number

2. COMPUTER SYSTEMS LABORATORY

3. Alan Demmerle
Acting Chief

I. SUMMARY

This Laboratory, now in its second year of operation, is currently pursuing three broad objectives in support of the conviction that computer technology can make substantial contributions to biomedical research in ways other than conventional data processing. These objectives are the development of processing systems for laboratory and clinical data, the development of data processing techniques, and the development of new applications for computers. The projects through which this Laboratory is pursuing these objectives include the design and development of computer systems to collect and manipulate data and control experiments in real time, systems for off-line data collection, a hybrid (analog/digital) computing facility, components of information retrieval systems, and other generally new applications of computers. All of these programs show great promise.

II. CURRENT LABORATORY PROGRAMS

The Computer Systems Laboratory, now in its second year of existence, has continued toward the system-oriented goals which it established last year. It has grown in size by about one third, and is now composed of 33 individuals representing the disciplines of engineering, mathematics, linguistics, and medicine, but who could be appropriately called Information System Specialists. Some of the projects started during the previous reporting period have carried over into this one; a few have been deferred, and many new ones have been undertaken.

A. Objectives

The objectives reported last year have not changed in substance, although the scope has been broadened. In order to see these objectives in a new perspective, they are set forth here in a different way from last year.

The Laboratory is dedicated to three broad objectives:

1. The development of processing systems for laboratory and clinical data

   a. Through the use of separate, dedicated, general-purpose computer systems

b. Through data collection systems for use in conjunction with a central computing facility

c. Through the use of special-purpose, hard-wired data processors.

2. Research, development, and implementation of new data processing techniques emphasizing the aspects of:

a. Methodology

b. Systems

c. Hardware

d. Software

3. Development and implementation of new applications for computers

The manner in which these objectives have been pursued through projects is summarized below.

B. Progress

The objectives previously set forth represent the current broad goals of this Laboratory. A number of projects have been undertaken in pursuit of each of these objectives. These projects, grouped under their general objectives, are briefly reviewed below. More details for many of the projects are included as Individual Project Reports.

Progress Toward Objective 1 (Development of processing systems for laboratory and clinical data)

There remain many project areas at NIH in which the use of a computer as a tool has not yet been explored or exploited. The laboratory and the clinical environment are particularly fertile areas in which computer technology can be applied. To date, however, our concentration has been primarily in the laboratory environment. It is our intention to help uncover project areas in both environments, and, by working in cooperation with the scientists involved, design systems to meet the data processing requirements of these project areas. The term "data processing" is to be considered in its broadest sense, and includes the collection, manipulation, computation, and display of data. These data processing requirements tend to fall into three categories:

a. Those requiring the full support of a general-purpose computer system at the site because of the size and complexity of the requirement or the number of requirements in a limited vicinity or the requirement for "real-time" processing of data and process control.

b. Those requiring data collection (and possibly display and transmission) for analysis, possibly at some later time, at some computing facility remote from the source of the data.

c. Those requiring specific processing which can be most economically implemented by hard-wired systems.

We have examined several project areas at NIH, determined the category prescribed by the requirements, and designed systems to meet these requirements.

1. a. We have designed, and are currently purchasing, a single computer system to serve all of the laboratories in NIDR jointly (described in detail in the Individual Project Report, DCRT 4.1). Similarly, we have designed, and are now purchasing, two other complete computer systems, one for some of the biochemical laboratories of NIAMD, and one for the Mass Spectrometer systems of LM, NHI (described in detail in the Individual Project Reports, DCRT 4.10 and 4.11). We have also designed three other complete computer systems, one to meet the needs of the NIMH scientists in Building 10 (DCRT 4.23), one for the Gerontology Branch of NICHHD, Baltimore (DCRT 4.24), and one for the Section on Comparative Behavior, NIMH, Poolesville (DCRT 4.25). These latter three systems, optimally designed to meet the specific needs of these groups, have not yet reached the procurement stage. All six systems are composed, in part, of commercially available equipment and software which will be purchased, and in part by non-available equipment and software which will be designed and implemented in this Laboratory. When implemented, these systems will represent a significant advance in the application of computation to biology and medicine. The systems will, in general, be used for collecting and analyzing data in real-time by having instruments directly connected to the computer; controlling experiments while they are being conducted; and developing new ways to do experiments which become possible only by virtue of the computational power of the computer.

In addition, of course, these systems will serve the traditional function of analyzing experimental data that has been collected. In addition to the system design, this Laboratory has assumed primary responsibility for achieving successful system operation. This involves developing the interface between the instruments and the computer, implementing remote communication links, and developing the applications and system controlling software. The systems will then, for the most part, be turned over to the scientists who expect to use them. In the final analysis, the value of these systems depends upon the laboratory scientists learning to use the computer as a tool since a fundamental appreciation of its capability is necessary if it is to continue to be used in an innovative way.

In addition to the actual design of computer systems for use at NIH, we have provided consultation on the design and use of computer systems for extramural programs, particularly to NHI and to NIMH. The largest

user of our expertise in this area was the NHI, whom we have helped on a continuing basis, particularly with the data management portion of the Myocardial Infarction Research Unit program. This is described in greater detail in Individual Project Report DCRT 4.12.

1. b. In some of the laboratories whose data processing requirements have been analyzed, complete computer systems are inappropriate at this time. In LM and LTD of NHI, and MET of NCI, we found that collection of data on digital magnetic tape, analog magnetic tape, and punched paper tape, respectively, and analysis of this data at a central facility was the best solution of their data processing requirements. These recommendations have not been acted upon to date. These cases are reported on in greater detail in the Individual Project Report DCRT 4.13.

For general use with such off-line data recording systems, we are developing time and event code generators and decoders. These allow the user of an automated data collection system to keep better track of the relationship between his experiment and his data. The generator will be used at the source of the data recording, and the decoder will be used at the site of the data processing. These units are described in greater detail in Individual Project Report DCRT 4.14.

1. c. Data processing requirements can sometimes be best met by special purpose computing systems, as opposed to the general purpose computers that we think of when we hear the word "computer." The special purpose systems, as the name implies, are designed for a specific job, and the logic is generally wired, although sometimes through a patchboard, so that they are not as easily adapted to other jobs as a stored program general purpose computer. They have a definite economic advantage when a single well-defined function is to be done repetitively. There are many uses for such machines at NIH. The monitoring of critically ill patients is one such area where these machines will be used in conjunction with general purpose machines. This Laboratory is currently developing such a special purpose computer for ECG analysis. The algorithms for the analysis were developed on this Laboratory's CDC 3100 computer system. Special purpose hardware, using these software algorithms, was then designed, and provides for a faster and more economical beat-to-beat ECG analysis than would be possible with a general purpose computer system. The end result is small, portable equipment for continuous real-time analysis of ECG data at the bedside. This is described in greater detail in the Individual Project Report DCRT 4.15.

<u>Progress Toward Objective 2</u> (Research, development, and implementation of new data processing techniques)

We have pursued this goal in four general areas.

a. Methodology - We are developing ideas relating to the organization of computing and memory systems. These ideas are quite imaginative, and their utility is under evaluation. They are described in the Individual Project Report DCRT 4.16.

b. Systems - The Hybrid Computer System (located in Building 10), after a fretful infancy, now offers hybrid computation, analog computation, digital computation, and analog to digital conversion. It is used currently nearly to capacity, and, hence, we are expanding its capability through the addition of a disk and a higher speed printer. To date, this system has been used mostly for standard digital computation rather than for hybrid computation. The hybrid computer provides a powerful tool for the solution of biomedical problems which involve biological systems modeling and simulation. The demand for this type of work has not met expectations, and problems requiring collaborative work in this area are being solicited. The work that has been done with this system is reported in some detail in the Individual Project Report DCRT 4.9.

c. Hardware - The nature of some data renders present data processing techniques uneconomical. As an example, data which, for short periods, are actively changing, but, for the rest of the time do not change, keep a computer system, or, at least, an input channel, busy all the time. We are developing "data compression" techniques to remedy this problem. These techniques will be implemented, probably with hardware external to the computer, in situations where the computer can be more productively used during periods of inactive data. This project is described in Individual Project Report DCRT 4.17.

d. Software - The processing, particularly searching and up-dating of large, active files of data, has, traditionally, led to long, complicated, and inefficient computer programs. New techniques leading to new software packages have been developed to minimize these problems, and are discussed in the Individual Project Report DCRT 4.18.

Progress Toward Objective 3 (Development and implementation of new applications for computers)

This objective, which overlaps Objective 1, due to its broadness, is an extremely important part of the mission of this Laboratory and Division. We have pursued this goal through several projects. We are currently working on a project which will lead to both the automatic abstracting of textual data and to better information retrieval systems. Key information, abstracted with the use of a computer doing linguistic analysis of autopsy reports, currently provides an input data source for a pathology data retrieval system. It will be extended for use in information retrieval systems using other types of medical textual data. The expectations and methods of this project are more thoroughly described in the Individual Project Report DCRT 4.19.

Another project which holds great promise is one to develop the methodology whereby the potential of the computer is made available in the physician's office at a reasonable cost. Such a system would permit access to the memory and computational abilities of a computer to relate symptoms to diseases and to compute treatment protocols for burn patients, etc. This fascinating application is discussed more fully in the Individual Project Report DCRT 4.20. Other applications of computation which have been pursued primarily through collaboration with other institutes include the Analysis of Laboratory Spectral Data, the Determination of Sequences of Protein and Nucleic Acids, the Analysis of Bone Density Measurements, and some work involving the small computer, called the Programmed Console, used in Radiation Therapy in the Clinical Center. These projects are reported on in the Individual Project Reports, DCRT 4.21, 4.5, 4.22, and 4.8, respectively.

General Remarks

In general, support available for the projects which have been discussed has been adequate, particularly with regard to finances, personnel resources, and enthusiasm from the management involved. However, space available to house personnel on the projects, and to house the electronics laboratory used in support of these projects has been totally inadequate. This has been detrimental to our programs. The members of this Laboratory look forward to the fulfillment of promises for more space.

There are a number of other projects, particularly in the clinical area, which could, and should, be undertaken, and which have the moral support of the management involved, but which cannot be undertaken now, or in the near future, without additional staff. Many projects in the computer system business cannot be undertaken halfheartedly since they require a substantial effort over an extended period of time to assure their success. Inadequate financial support and personnel would undoubtedly lead to failure of these projects.

Serial No. DCRT 4.1
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Dental Institute Computing System

Previous Serial Number: DCRT 4.1

Principal Investigator: Daniel Syed

Other Investigators: William Holsinger

Cooperating Units: National Institute of Dental Research

Man Years:

Total: .5
Professional: .5
Other: 0

Project Description:

Objectives:

This project is intended to provide on-line real-time computational support for the biochemistry and neurophysiology research programs of the National Institute of Dental Research. The computer system is designed to automate data acquisition from instruments in the biochemistry laboratories, to perform limited computations on captured data, and eventually to control experiments. During neurophysiological research procedures, the system will provide information relative to the proper placement of electrodes, pertinent waveforms enhanced by averaging techniques to improve signal-to-noise ratios, and correlative relationships. The system will provide a test bed for the development of hardware and software techniques for automating many instruments that have not been automated on even an off-line basis.

Methods:

The analysis and system specifications were submitted through departmental channels to obtain data processing clearance. Considerable delay was encountered due to DHEW reorganization. Program approval was received in February, at which time a request for proposal was released. Responses to this request for proposal are being evaluated.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

The proposed system will permit this Division to develop software and hardware for use in similar applications around the campus. Such a system will be used as a research tool in the design of an optimal system for biomedical applications stressing data acquisition. Consistent with the development of time-sharing software on the Central Computer Facility IBM 360/50, the system will provide a basis for analysis and development of satellite computing techniques in a biomedical environment.

Proposed Course:

Upon selection of a specific manufacturer and negotiation of an acceptable contract, special purpose hardware for interfacing the instruments with the selected computer will be developed. Associated software for support of research programs in real-time, including substantial modifications to a real time monitor, will be developed concurrently.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.3
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Study of Hybrid Computer Utilization in Support of Laboratory of Metabolism, NHI

Previous Serial Number: DCRT 4.3

Principal Investigator: Daniel Syed

Other Investigators: William Holsinger

Cooperating Units: None

Man Years:

Total: 0
Professional: 0
Other: 0

Project Description:

Objectives:

This study was designed to investigate the feasibility of expanding the current hybrid configuration to support biochemistry research programs of the Laboratory of Metabolism, National Heart Institute. Since the cost of modification of the hybrid facility was considered prohibitive and the alternate methods were not considered acceptable, the project was terminated.

Methods Employed: None

Major Findings: None

Significance to Biomedical Research and the Program of DCRT: None

Proposed Course: None

Honors and Awards: None

Publications: None

Serial No. DCRT 4.4
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Hospital Patient Record and Information and Retrieval System

Previous Serial Number: DCRT 4.4

Principal Investigator: Victor Colburn

Other Investigators: William White, Michael Otten

Cooperating Units: None

Man Years:

Total: 1.2
Professional: 1.2
Other: 0

Project Description:

Objectives:

Early in fiscal 1968, the previously reported objectives of this project were shifted to include procurement and implementation of a general-purpose computer display facility. The previous objectives involved a planned approach to application of computer system technology within the NIH Clinical Center. These earlier objectives remain intact, with the intention that they will constitute a primary use of the proposed new Computer Display Facility.

Methods Employed:

The major activity on this project during fiscal 1968 consisted of the preparation of a functional specification suitable for normal RFP procedures.

Major Findings:

A specification for a real-time Computer Display Facility was completed at about mid-year fiscal 1968. The specification provides for a multi-user time-shared facility with up to 66 remote terminals, including cathode-ray tube displays of both character type and the more sophisticated graphic varieties. Terminals will contain keyboards and machine readers for control and data inputs. There is

provision for automatic input of data from bio-medical and laboratory instrumentation. The facility will include a mass storage system. Strong emphasis is placed on interactive operation and user convenience.

Further effort was curtailed at this point, due primarily to budget cuts occurring at about midyear fiscal 1968, coupled with lack of a hard requirement for such a computer facility, for example, strong justification from an existing project outside of DCRT. However, portions of the above-mentioned specifications are being used for procurement of a computer system for Pharmacology and Toxicology, in a collaborative effort between DRFR and DCRT.

There is little purpose in the further pursuit of the originally stated objectives of this project, until procurement of such a computer display facility can be reactivated.

Significance to Biomedical Research and the Program of DCRT:

The procurement by DCRT of the proposed general purpose real-time computer display facility still appears to be a worthy objective. We are convinced that when such a facility is available, it will find much application. Many computer applications now implemented on teletype remote service will find new life in the much greater speed and flexibility of the proposed facility.

Proposed Course of Project:

The project is currently dormant due to lack of funds. If adequate funds can be allocated within the next 1 to 2 year period, it should probably be reactivated, without significant change.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.5
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Sequence Determination of Proteins and Nucleic Acids

Previous Serial Number: 4.5

Principal Investigator: Marvin Shapiro

Other Investigators: Jay Vinton, Lewis Sheiner, Carl Merril

Cooperating Units: Laboratory of Neurochemistry, NIMH; Laboratory of General and Comparative Biochemistry, NIMH

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

(1) The mathematical reconstruction of protein and RNA sequences from laboratory data and (2) the computer simulation of the laboratory process of sequencing a biopolymer.

Methods Employed:

The development of an algorithm for reconstructing sequences and the use of this algorithm in a program which simulated the steps followed in the laboratory sequencing of an RNA molecule.

Major Findings:

The simulation program was completed. Different sets of laboratory conditions were postulated (including sequence size and composition, amount and types of equipment available, and laboratory personnel available) and computer runs were made.

Significance to Biomedical Research and the Program of DCRT:

(1) The sequencing process was formalized.
(2) Computer results indicated bottlenecks in the process due to

lack of essential equipment or personnel and gave estimates of the time needed to sequence RNA polymers of prescribed lengths and compositions.

Proposed Course:

Further computer runs will be made to obtain further information and statistics in order to gain insight into the sequencing problem. Possible modifications will be made in the program to make the simulation more accurate.

Honors and Awards: None

Publications:

Shapiro, M. B.: An Algorithm for Reconstructing Protein and RNA Sequences. J. Assoc. Computing Mach. 14, 720-731, October, 1967

Serial No. DCRT 4.8
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Support of Programmed Console

Previous Serial Number: 4.8

Principal Investigator: William White

Other Investigators: None

Cooperating Units: Radiation Branch, NCI

Man Years:

Professional: .2
Other: 0
Total: .2

Project Description:

Objectives:

Continuing support for the use of the programmed console has included consultation, suggestions for hardware addition, and the programming of an assembler for the program console symbolic language.

Methods Employed:

A program written in PL/1 for the 360/50 accepts the symbolic machine language of the Spear PC, and produces a bit pattern in 12 bit words which is an "object program" for the Programmed Console. The PL/1 program produces a listing containing the address in decimal and octal, and the instruction in binary, octal, and symbolic.

Major Findings:

The assembler appears to be a useful addition to the software of the Programmed Console. The experience of writing this type of program in PL/1 demonstrates the utility of the PL/1 language, and suggests other similar conversion applications.

Significance to Biomedical Research and the Program of DCRT:

The use of small special-purpose computers in research projects and patient care situations offers the advantages of immediate access, fast response, and the ability of the investigator to interact with his program and the computer. The limitations of size, storage capability, and versatility of the small computer can be surmounted by support from, and communication with, the larger NIH computer facility.

Proposed Course:

It is proposed that a direct communication link from the 360/50 system to the Programmed Console be established as soon as feasible. When this is accomplished, the assembler program will have the capability of loading a program directly into the Programmed Console. Other support activities will involve the programming for cataloging and storing programs and data from the Programmed Console in the 360/50 system and the development of additional software such as a FORTRAN compiler for the Programmed Console.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.9
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Hybrid Computer Project

Previous Serial Number: Same

Principal Investigator: Perry S. Plexico

Other Investigators: Leroy Brown, S. J. Farlow, Robert Romanoff, David Songco, Philip Turner

Cooperating Units: Section on Clinical Biophysics, Cardiology Branch, NHI

Man Years:

Total: 4.7
Professional: 3.7
Other: 1

Project Description:

Objectives:

1. To develop a hybrid analog/digital computer system and methods of applying it to the simulation and modeling of biological systems.

2. To provide methods and capability for data conversion, editing, and real-time analysis.

Methods Employed:

Mathematical techniques for the optimization and correlation of models of biological systems to experimental data have been developed, and computer programs written to execute these techniques on the hybrid computer. Software was completed to accomplish real-time data conversion and editing, and to allow the use of FORTRAN for certain classes of real-time data analysis.

Major Findings:

Efforts to correct the serious engineering design problems which existed in the analog computer and linkage portions of the hybrid system were continued, with the result that the system was brought

to full operational status in August 1967. Sample-and-hold amplifiers to be used as an aid to data conversion were designed and completed in May, 1968.

As a test of the hybrid optimization techniques which were developed, a number of test problems were executed. These same problems were repeated using a strictly digital computer method, wherein the analog computer was digitally simulated. Results of these tests were in close agreement, but the hybrid implementation had a speed advantage of several thousand to one over the digital method. Following these tests, a mathematical model of cardiac muscle was simulated on the hybrid computer, and an attempt made to correlate it with experimental data using a gradient optimization procedure. While a reasonable correlation was not obtained, the detailed study of the model made necessary during the optimization attempt revealed several features in its mathematical formulation which violated the original hypotheses on which it was based.

Significance to Biomedical Research and the Program of DCRT:

1. The hybrid computer was shown to be of significant value for the simulation and modeling of biological systems which may be mathematically expressed in terms of differential equations.

2. A number of investigators from several Institutes and Divisions have been informally trained in the use of the hybrid system and have applied it to a variety of scientific computation tasks.

3. At this time, the hybrid computer offers the only data conversion and real-time data analysis facility available on an NIH-wide basis. The capability of applying the FORTRAN programming language for certain of these functions greatly simplifies its use.

Proposed Course of Project:

1. Efforts of this project will continue to be directed to the simulation and modeling of biological systems. Problems currently under consideration include in vivo studies of pulsatile blood flow, fluid flow through collapsible tubes (of significance in both pulmonary and vascular mechanics), and optimal muscle control.

2. The digital portion of the system is to be expanded to include mass data storage capability and a higher speed printer which will substantially improve the system's operating efficiency. It is the intent of this project to further utilize this increased capability to perform research in multi-programming and time-sharing, hopefully including the processing of real-time laboratory data from remote locations.

Honors and Awards: None

Publications:

Plexico, Perry S.: Multiparameter Optimization of Dynamic Systems Using a Hybrid Computer. M.S. Thesis, George Washington University, February, 1968.

Serial No. DCRT - 4.10
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Computer System for Biochemical Laboratories of NIAMD

Previous Serial Number: None

Principal Investigator: Marvin Shapiro

Other Investigators: Jeffrey Buzen, Isaac Hantman, Kenneth Kempner, Arthur Schultz, Jay Vinton

Cooperating Units: Laboratories of Physical Biology and Molecular Biology, NIAMD

Man Years:

Total: 2
Professional: 2
Other: 0

Project Description:

Objectives:

To develop a real-time data acquisition and analysis computer system to service the laboratories in Building 2, NIAMD.

Methods Employed:

The computational needs of a representative group of scientists in Building 2 were examined from the standpoint of equipment used, data rates, methods of data analysis, and possible application of feedback methods.

Major Findings:

Specifications were written for a real-time computer system for Building 2, and a request for proposals was prepared.

Significance to Biomedical Research and the Program of DCRT:

Implementation of the proposed computer system would have far-reaching implications: (1) Present turn-around time for computer results

in Building 2 would change from days to minutes. (2) Close interaction between scientist and computer would make new methods of data analysis possible. (3) Real-time data collection would give more accurate results than are now obtainable. (4) The on-line system would be far more flexible and accessible than the current batch processing use of the remotely located IBM 360/50 computer.

Proposed Course of Project:

A computer will be selected, and programs will be written to accomplish acquisition and analysis of data from instruments interfaced directly with the computer.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.11
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Computer Applications to Mass Spectrometer Data

Previous Serial Number: None

Principal Investigator: Eli Gilbert

Other Investigators: Henry Fales, George W. Milne, Jay Vinton

Cooperating Units: Laboratory of Metabolism, NHI

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

To develop a computer system for reducing data produced by a mass spectrometer and computer programs for further analyzing the data to give insights into the topological structure of complex molecules.

Methods Employed:

Specifications were developed for a computer system to be directly connected to Dr. Fales' mass spectrometers. The system will accept data in real-time and perform some data reduction. Programs were completed for the IBM 360/50 and Hybrid computers, to handle (1) digitization of raw data (2) recognition of peaks (3) standardization of peaks and (4) calculation, for each peak, of a centroid, area, and empirical molecular formula. In addition, plotter routines were written for displaying the possible paths followed by the molecule in fragmenting.

Major Findings:

A rapid turnaround system was implemented for mass spectrometer data. Data so reduced will be used as the basis for the application of heuristic methods in molecular structure determinations.

Significance to Biomedical Research and the Program of DCRT:

Through use of the computer, a significant increase in the amount of mass spectrometer data analyzed was realized. In addition, some of the computer output, notably the plotter results, provided new insights into the underlying molecular structures.

Proposed Course of Project:

(1) Analysis programs will be completed. (2) A large effort has begun toward applying heuristic methods to structure determination and (3) a small real-time computer system has been requested, to be directly connected to the mass spectrometer. Its implementation will increase the overall data throughput rate of the current system by an order of magnitude.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.12
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Data Management Systems for the Myocardial Infarction Research Unit Program

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigators: Kenneth Kempner

Cooperating Units: None

Man Years:

Total: .25
Professional: .25
Other: 0

Project Description:

Objectives:

Five institutions have been funded by the National Heart Institute to study the causes and characteristics of myocardial infarction. This research program, it is hoped, will lead to improved techniques in both the prevention and treatment of this disease. This project is intended to provide guidance in the establishment of individual data management systems to the five myocardial infarction research units already funded, and to the seven additional institutions scheduled to receive funds. A primary goal is to promote compatibility of hardware and software to the maximum extent possible. A secondary objective is the assimilation of "state of the art" techniques in cardiac monitoring in preparation for the design of a computer system for the recovery area of the Clinical Center heart surgical suite.

Methods Employed:

Guidance to the myocardial infarction research units is accomplished primarily through consultation provided to the National Heart Institute. Although the computation and data management plans of each institution involved have been reviewed and have frequently resulted in major recommendations, a most important facet of the project has

been the opportunity to visit and evaluate personally several leading non-MIRU institutions engaged in cardiovascular research, all the MIRU institutions thus far funded, and also many of the leading computer manufacturers that rank as potential MIRU contractors. It is this facet, in fact, which has made it possible for DCRT to offer some tangible guidelines and goals to the National Heart Institute, independent of submissions from the selected myocardial infarction research units.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

This project provides a basis for evaluating the feasibility of attempting collaboration between institutions establishing individual data management systems directed toward a similar goal. It will also provide experience in cardiac monitoring and care. This experience will be invaluable in prosecuting projects within the heart surgical suite in the Clinical Center.

Proposed Course:

Consultation to the National Heart Institute in the area of the establishment of data management systems will continue. Eventually, this consultation will be extended to include guidance in the analysis of data acquired by the various institutions involved.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.13
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Special Purpose Data Processing Systems for NHI and NCI

Previous Serial Number: None

Principal Investigator: Paul Heffner

Other Investigators: David Songco

Cooperating Units: (1) Laboratory of Metabolism, NHI
(2) Laboratory of Technical Development, NHI
(3) Metabolism Branch, NCI

Man Years:

Professional: .75
Other: 0
Total: .75

Project Description:

Objectives:

To analyze laboratory data collection requirements, and design the appropriate system to collect and prepare the data from these laboratories for computer analysis.

Methods Employed:

During this reporting period, the computing requirements of three laboratories were analyzed. The most effective and economical approach to these requirements was through off-line collection of data which is then manually transported to a computer system. Each system is summarized below.

1) Laboratory of Metabolism, NHI
There are several gas and liquid chromatograph instruments in this laboratory, the data from which requires analysis. This data can be best collected on a unified data storage system whereby the data from each instrument would be multiplexed on to a single digital magnetic tape. Once a day this tape could be taken to a computer center where it would be processed, and the results available to the experimenters in less than 24 hours from the initiation of the experiment. The computer would

calculate the corrected areas under peaks, and locate the centers of the peaks with respect to calibration runs.

2) Laboratory of Technical Development, NHI
Several particle-detecting instruments are used to make measurements of the growth of organisms which are captured in an agar media. A data storage medium is required to collect data from the above experiment and to transport it to a computer system. Software is required to calculate pulse height and pulse width distributions associated with the detection of the particles. An analog tape recording system was recommended.

3) Metabolism Branch, NCI
Five liquid and crystal scintillation systems produce data which is currently analyzed manually. A paper tape system for data collection, formatting, and storage would allow for statistical analysis of this data on a computer system. Programming is required to perform the analysis and to present the results in printed form and on punched cards.

Major Findings:

The advantage common to each system is that it reduces to a minimum the tedious and time-consuming hand calculations that are presently performed in processing the raw data. Secondly, the accuracies that will be afforded by the computerized techniques will exceed those of hand calculation.

Significance to Biomedical Research and the Program of DCRT:

In reducing the time required by the researcher to process his data, more time will be available to him to extend his investigations. Further, the computerized analyses can provide additional information about the data through algorithms too involved for repetitious hand calculations, and can remove deleterious data from calculations.

Proposed Course of Project:

If these systems are acceptable to the management of the laboratories concerned, CSL could oversee their implementation.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.14
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Combined Event and Time Code Generator and Reader Units

Previous Serial Number: None

Principal Investigators: Paul Heffner - David Songco

Other Investigators: None

Cooperating Units: None

Man Years:

Total: .5
Professional: .5
Other: 0

Project Description:

Objectives:

The objective of this project is to make a time and event code generator available to NIH experimenters who record their test data on analog tapes. By combining both an event encoder and decoder and a precision time code generator and reader into one compact unit, an optimal capability for the marking of recorded experimental data will be effected.

Methods Employed:

Functionally, the event code generator will place a manually programmed identification code on the carrier of a modified IRIG B BCD time code, in addition to the time code, so that the time code continuity will not be disrupted when an event is marked. Time will be resolved in seconds, minutes, and hours, and will be accurate to better than .002%. The day of year count can, in addition, be placed in the code if the operator so chooses. The decoding capability of the unit will enable an experimenter to play back his data from his recorded magnetic tape and to relocate a pre-programmed part of his experiment. In addition, the unit will be made compatible with the CDC 3100 computer in Building 10 for automatic location of experimental data during the processing of recorded data, and to enable computer entry of the absolute time upon computer request. The system has been designed to employ

integrated circuits primarily.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

The availability of a precision time code signal during the recording of experimental data will enable the experimenter to reconstitute time along with his data during later processing of the recorded experiment. The signal will also enable an experimenter to evaluate his own recording and reproducing process with regard to both dynamic and static playback tape speed errors. For archiving purposes, the time code can log the time and day of year of the test as a permanent record on the tape.

Proposed Course:

The first four units are expected to be completed by the Fall of 1968. If more units are desired by experimenters, the Computer Systems Laboratory can oversee their fabrication.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.15
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Beat-to-Beat ECG Analysis

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigators: Kenneth Kempner, Martin Miller, William Holsinger

Cooperating Units: None

Man Years:

Total: .75
Professional: .75
Other: 0

Project Description:

Objectives:

The primary objective of this research program is the development of real time ECG rhythm analysis techniques for the detection and classification of cardiac arrhythmias. Milestones in this program are the identification of parameters which will differentiate between the most commonly occurring arrhythmias and the development of computer software and hardware systems for the acquisition and reduction of the ECG waveform.

Methods Employed:

In the area of computer software, several programs have been developed which allow the detection of R waves, computation of R-R interval Histograms, and the calculation of the correlation coefficient and RMS error between two ECG complexes. These computations are accomplished in better than real time using a data set of representative arrhythmias stored on magnetic tape. The criteria developed for the detection of the QRS complex is insensitive to the presence of baseline drift and noise.

A special purpose digital system has been designed as the hardware counterpart of the software system. This equipment is expected to perform the arrhythmia detection and classification without the support of a general purpose, stored program digital computer.

Major Findings:

It was found that correlation coefficients and RMS errors provide a valid indication of the dissimilarity between two ECG waveforms.

Significance to Biomedical Research and the Program of DCRT:

This project is applicable to the long term monitoring problem in intensive care units and directly supports the MIRU program.

Proposed Course of Project:

The software phase of this project will continue with the development of programs to investigate the frequency content of cardiac arrhythmias as a diagnostic factor. The addition of real-time data acquisition software is planned and will permit on-line acquisition and A/D conversion of electrocardiograms. Finally, analysis of electrocardiograms from remotely located patients will be performed utilizing a standard telephone transmission link or equivalent. Concurrently with the software development, the hardware digital arrhythmia analyzer will be tested, evaluated and modified, as appropriate, in order to optimize the accuracy and scope of the system.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.16
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Principles of Global Organization in Dense Volumetric Memories

Previous Serial Number: None

Principal Investigator: Harry Blum

Other Investigators: None

Man Years:

Total: .5
Professional: .5
Other: 0

Cooperating Units: None

Project Description:

Objectives:

To expand our conceptual base for understanding intellectual operation in animals and machines by providing a theoretical foundation for the parallel, pre-attentive, pre-logical operations.

Methods Employed:

Formalizing, developing, exploring and simulating models of parallel information processing in dense volumetric memory structures. These models use propagative pulse processes to facilitate a complete interchange of information widely distributed in memory. The project uses insights from widely diverse disciplines--mathematics, engineering, physiology, and psychology, etc.--to explore both structure and function in this ill-understood area.

Major Findings:

Formal physical models of such parallel processes have been previously defined. They have been extended this year to define rudimentary information structuring operations. A simple perceptual type "grammar" has been defined as a first step toward formalizing a method of imaging

a world by small incremental perceptual operations that can be carried out by a parallel interactive processor. Relationships with perceptual processes and global brain function have been extended.

Significance to Biomedical Research and the Program of DCRT:

The work has broad potential in extending experimental directions in psychology and neurophysiology by proposing principles and constraints that may be operative in (1) the pre-attentive perceptual processes and (2) global brain processes. If relevant, the latter has important relationships with EEG and the non-specific projection processes such as are encountered in the recruiting and arousal phenomena of neuropsychology. In addition, the work is useful to DCRT and the general biomedical field, by proposing information retrieval structures which are geared to the imaging of a rich multi-dimensional cognitive, semantic and linguistic world. These structures have potential applicability to the study of such processes in higher animals and in developing software and hardware for accomplishing the functions with mechanism.

Proposed Course of Project:

1. To develop formal methods and bases for parallel intuitive processes.

2. To develop collaborative efforts in experimental psychology and neurophysiology within NIH and/or outside.

3. Simulate and explore, on digital computer, semantic structures and information retrieval problems occurring in a rich world.

4. To study the problem of new devices.

Honors and Awards: None

Publications:

Blum, Harry: "Global Brain Function: A Model Using Autonomous Broadcast Elements," Proceedings of Technical and Biological Problems of Control Symposium, Yerivan, Armenia, USSR Sept. 1968

Serial No. DCRT 4.17
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Data Compaction

Previous Serial Number: None

Principal Investigator: Paul Heffner

Other Investigators: None

Cooperation Units: None

Man Years:

Total: .1
Professional: .1
Other: 0

Project Description:

Objectives:

The objectives of the data compaction project are to reduce memory requirements when voluminous data require storage, and to enable data that would otherwise overload input/output channels to be transmitted and received through the channels.

Methods Employed:

Study has thus far centered around compression techniques that can be applied to continuously occurring analog data that is sampled periodically and stored for later reconstruction and processing. The techniques involved are those which could be implemented either with a computer, or, if appropriate, with external hardware.

Major Findings:

Two basic techniques were found which, for most analog signals, would allow implementation through straightforward software or hardware design. The first basic technique is that referred to as a zero order system extrapolation and reconstruction technique, and the second as a first order system extrapolation and reconstruction technique. The degree to which both of these techniques will provide storage reduction

has been shown to depend upon the statistical nature of the data signal, the sampling rate, and the margin of error allowed by the experimenter in the reconstruction process. Algorithms for both techniques were generated.

Significance to Biomedical Research and the Program of DCRT:

One of the limiting factors facing on-line data acquisition systems is storage capacity for continuous high data volume experiments. Implementation of the above generated algorithms, either by software or by hardware, will allow continuous storage of data into one storage medium without causing data discontinuity. Some biological data requires fast sampling rates because it is very active and rapidly changing part of the time, even though it is inactive the rest of the time. If this data were to be captured and analyzed in a time-shared computer system, best use could be made of the system if the channel through which that data flows were allowed to be inactive when the data were inactive. A hardware data compressor, external to the computer, would allow for this type of system optimization.

Proposed Course of Project:

These techniques will be implemented for a specific application.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.18
1. Computer Systems Laboratory
2. Not applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: File Structure and Retrieval Processing System

Previous Serial Number: None

Principal Investigators: Scott I. Allen and William C. White

Other Investigators: John K. Knight and Thomas L. Tubbesing

Cooperating Units: None

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

Most information storage and retrieval projects appear to have special requirements that preclude the use of an all-inclusive general purpose information processing language; however, the creation of a user-oriented set of well specified data processing routines can greatly simplify the programming tasks of file creation, and update data retrieval and report generation. The purpose of this project is the production and testing of such a package of routines to handle complex moderately large research data bases with a minimum of programming effort.

Methods Employed:

Preliminary experience at DCRT suggests that the programming language Pl/1 offers a sophisticated new tool which is relatively easy to learn and use for non-numeric computational data processing. This project uses PL/1 as the primary programming language to construct subroutines and macro instructions to be included in a user's program tailored to his specific needs.

The software capabilities include brief standard job control language sets, sort/merge commands for the organization of retrieval data, a PL/1 callable sort/merge subroutine, convenient plotter and line printer

graph producing capabilities, powerful table-making features and a date conversion routine.

Major Findings:

Preliminary tests on a small synthetic data base have shown that the concept and implementation of a PL/1 language set of sub-routines and macros is a workable system for information retrieval problems. The system is being tested on an actual medical information system for approximately 150 patients.

Significance to Biomedical Research and the Program of DCRT:

This system of file structure and retrieval processing programs should be useful for many medical research data files. It will enable the DCRT programmers to construct data storage and retrieval programs with a minimum of effort.

Proposed Course of Project:

After extensive testing of the system, it is planned to make a program description manual available for users. This project should be completed in the present fiscal year.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.19
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Information Retrieval and Natural Language Data Processing

Previous Serial Number: None

Principal Investigator: Milos Pacak

Other Investigators: George Dunham, Helen DeFrancesco, Martin Epstein

Cooperating Units: None

Man Years:

Total: 3.5
Professional: 3.5
Other: 0

Project Description:

Objectives:

The goal is the development of a system for meaningful and timely retrieval of medical data for use by the scientific community.

a. Evaluation and implementation of modern automated data handling techniques and methods of linguistic analysis in order to handle narrative medical text in the pursuit of the overall goal.

b. The development of new techniques for programming, linguistic analysis and retrieval which are responsive to the user, that is, the medical community.

In the first stage, the research was limited to the early work in a manageable medical specialty, which represents a cross section of data handling and retrieval problems in the field of medicine.

Methods Employed:

The natural language data (NLD) processing system is represented by the 3-tuple

$$NLD = (T, D, G)$$

where T is the input text (pathology reports), D denotes the dictionary (SNOP) and G denotes context-sensitive grammar.

Grammatical analysis is performed on 3 levels:

a. Morphological analysis which implies the identification of productive terminal morphemes as markers of part-of-speech classes, and the set of transformation rules by which canonical forms are derived from adjectives and nouns. Nouns or their substitutes constitute the key semantic elements in the system and must often be derived from adjectives or other nominals.

b. Syntactic analysis implies the recognition of pseudo-sentence (utterances) boundaries, and the recognition of boundaries of noun phrases which constitute the conceptual units in the information retrieval system.

c. Semantic analysis is focused on the establishment of semantic correlations among the elements belonging to one of four major semantic categories as they are listed in the SNOP dictionary.

Computer implementation of morphological analysis is well underway. Study of the currently available programming languages and techniques for use in the project is a continuing effort.

Early and continual involvement of the user insures responsiveness of the system and allows for the evaluation of its effectiveness and acceptability. The retrieval needs of the scientific community are being carefully examined.

Major Findings:

1. The medical subfield of Pathology has been representative of problems across the field of medicine. A well-constructed dictionary of pathological terms, SNOP, was already in existence and has provided an excellent semantic and conceptual dictionary for the implementation of our techniques.

2. Medical text proved to be very useful for linguistic analysis.

3. The interest and support of the user has been stimulated, and it is anticipated that the system will represent a major contribution to the medical community.

Significance to Biomedical Research and the Program of DCRT:

The majority of medical data are recorded in natural language form (English language). In any linguistic analysis, it is necessary to assume that words convey certain meaning and that their meaning is

determinable. The relation between a word and its meaning is the semantic correlation which is established by agreement and which, by acceptance, becor s convention. It is recognized today that computational linguistics is not only useful but necessary in any field in which the computer is called to interpret or analyze natural language, that is, indexing and classifying documents, indexing or abstracting pertinent documents, preparing computer-oriented microglossaries, analyzing the language of speakers with mental disorders, etc. The analysis and retrieval of medical data recorded in natural English by a computer requires the implementation of sophisticated techniques for natural language data processing.

Proposed Course:

Continuous refinement of the program with the main emphasis on the semantic analysis.

Honors and Awards: None

Publications:

Pacak, Milos: "Computational Morphology" in Austin, W. (Ed): Papers in Linguistics in Honor of Leon Dostert, Mouton Co., The Hague, 1967, pp. 134-145.

Pacak, Milos: "Homographs: Classification and Identification" in Etudes de Linguistique Applique (Studies in Applied Linguistics), Didier, Paris, pp. 89-105, 1968.

Serial No. DCRT - 4.20
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Experimental Medical Telecommunications

Previous Serial Number: None

Principal Investigator: Scott I. Allen, M.D.

Other Investigators: Michael Otten, Eric Swarthe, William White

Cooperating Units: None

Man Years:

Total:	1.3
Professional:	1.3
Other:	0

Project Description:

Objectives:

This project is concerned with the design, implementation, and field testing of new computer concepts and technology, as applied to medical information communication. It is oriented toward the goal of the continuing education of health professionals by providing rapid access to critical facts and calculations needed in patient-care decision-making by means of simple telephone communication facilities. The basic purpose of the project is to provide a method of giving the practicing physician and professional associates low-cost, easy-to-use, fast access to current medical information, computational assistance in the therapy protocols; and diagnostic assistance by computer programs used on a shared basis.

Methods Employed:

The basic terminal device being investigated in this project is the standard telephone available in every doctor's office. The lowest cost service can be supplied using a standard or Touch-tone telephone for the input of requests in digital codes, and the same telephone to receive an audio computer-generated output response. Other low-cost hard copy output devices are being investigated. Prototype programs have been written which include pediatric burn treatment calculation, intravenous drug compatibility information, disease-symptom complex

listings for diagnostic assistance and a computer-aided instruction example of inter-active teaching dialogues. Computer service is supplied by commercial time-sharing systems, and is the most economical method of demonstrating the feasibility of the project. A Transmission Control Computer to handle line-switching, message buffering, code translating, and analog conversion is planned as the link between the telephone equipment in the doctor's office and the large, time-shared computers.

Major Findings:

Problems have been identified and solved in regard to the compatibility, reliability and availability of telephone communication equipment. Actual demonstrations have shown the project concepts to be economically feasible and have excited considerable interest in the project from medical administrators.

Significance to Biomedical Research and the Program of DCRT:

The technology being developed can be used to make research information readily available to the research scientist and medical practitioners. The ease of use, low-cost and ready acceptability of telephone input and audio response are receiving consideration in future plans for the main computational facility at NIH as provided by DCRT. The programming techniques and technical requirements for time-sharing, message switching and analog signal processing for communication by voice grade telephone lines can be of use in future expansion of the system facilities of DCRT and other university medical centers.

Proposed Course of Project:

The proposed course of this project envisages collaboration between NIH and selected university medical centers in various forms. Activities such as data base construction and maintenance, program specification and implementation, medical office field tests, service evaluations and equipment engineering could be sponsored by university medical centers in collaboration with DCRT. Prototype computer systems and devices would be tested and evaluated at DCRT and medical centers. Recommendations and specifications for the implementation of biomedical information networks would be prepared.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.21
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Analysis of Laboratory Spectral Data

Previous Serial Number: None

Principal Investigator: Marvin Shapiro

Other Investigators: Marie Chang, Carl Merril, H. Todd Miles

Cooperating Units: Laboratory of Neurochemistry, NIMH; Laboratory of Molecular Biology, NIAMD

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

Mathematical curve fitting and pattern recognition methods are being implemented on the computer to recognize data patterns and to fit theoretical functions to the data.

Major Findings:

Programs have been written for analyzing ultraviolet (UV) and infrared (IR) spectra of oligonucleotides and the deconvoluting a series of overlapping data peaks.

Significance to Biomedical Research and the Program of DCRT:

The computer programs described should eventually have wide applicability at NIH, both by the many laboratories employing the same instruments for which the present programs were written, and by laboratories with other similar types of data which needs deconvoluting.

Proposed Course of Project:

The programs are in a preliminary stage of development. Work is being done in the areas of improving the methods of numerical analysis and curve fitting, and in setting up files of standardized data to be used for matching and curve-fitting. Eventually, many of the techniques and programs being developed will be used with on-line application, such as in NIAMD, Building 2.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.22
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Bone Mineral Mass Measurement in-vivo

Previous Serial Number: None

Principal Investigator: William C. White

Other Investigators: William Vincent

Cooperating Units: Clinical Endocrinology Branch, NHI

Man Years:

Total: .25
Professional: .25
Other: 0

Project Description:

Objectives:

The purpose of the clinical study pursued by Dr. Ronald G. Evens of the Heart Institute is the detection of diseases which are characterized by a loss of mineral content of the bone. Quantitation of mineral loss has been a difficult clinical problem. A method of measurement of bone density using a mono-energetic gamma ray source to measure the absorption of the bone in-vivo has been developed. The calculations required to determine the integral of the absorption over the bone area studied are fundamentally simple, but variations in the measured data necessitates some smoothing techniques.

Methods Employed:

The data from the bone density measurements were recorded on punched paper tape, which was converted to punched cards, for input to a computer program written in FORTRAN and run on the 360/50. Reliability of the calculations and reproducibility of the measurements were checked by use of standard samples and multiple measurements.

Major Findings:

The bone density measurement method with computer analysis of the data has been shown to be a reproducible estimation of bone mineral content and useful for diagnosis and quantification of bone changes.

Significance to Biomedical Research and the Program of DCRT:

The calculations used in this FORTRAN program are simple and direct, and do not involve any new mathematical techniques. The program can be useful in other photon absorption problems.

Proposed Course of Project:

The project is being continued on a routine service basis.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.23
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Remote and Satellite Computation for National Institute of Mental Health Building 10 Scientists

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigators: William Holsinger and Martin Miller

Cooperating Units: None

Man Years:

Total: 3.5
Professional: 3.5
Other: 0

Project Description:

Objectives:

This project is intended to provide both on-line real-time and off-line computational support for research programs of National Institute of Mental Health Scientists currently scheduled to remain in Building 10. Specifically, on-line real-time data acquisition and computation is proposed in support of experiments in Problem Solving, Learning, Sleep and Dreaming, and Perception, and also in support of experiments relating the analysis of electroencephalographs and physiological variables to clinical diagnosis and classification of NIMH patients. Off-line computation is proposed in support of studies involving voluminous statistical processing as in the analysis of data from various surveys.

Methods Employed:

The on-line real-time requirements of Building 10 NIMH scientists were studied, and a local computer system specified. A similar analysis of off-line requirements led to the specification of a remote terminal to be located within NIMH Building 10 facilities.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

Consistent with the development of time-sharing software on the Central Computer Facility IBM 360/50, the system will provide a basis for analysis and development of satellite computing techniques in a biomedical environment characterized by the presence of both high volume off-line data processing requirements and sophisticated on-line real-time requirements.

Proposed Course of Project:

Upon receipt of program approval, the satellite computing system will be procured on open bid. System software will be generated for this system. A remote terminal consisting basically of a small IBM 360/20 and associated printer and card equipment will be installed at NIMH in the next six months. This remote facility will operate in a queued batch process mode, and will utilize DCRT developed software that is presently being tested out.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.24
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Gerontology Research Center Computing System

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigator: Isaac Hantman

Cooperating Units: None

Man Years:

Total: 2
Professional: 2
Other: 0

Project Description:

Objectives:

This project is intended to provide support for research programs of the Gerontology Research Center in the areas of on-line computation and control, analog to digital conversion and off-line data processing. The on-line computation capability is designed to support experiments in Problem Solving while the on-line control capability is to be utilized in the control of blood glucose/insulin infusion rates. The analog to digital conversion and off-line computation capabilities will be used in the analysis of electrocardiograms, physiological data, scintillation counter data and nuclear magnetic resonance data, etc.

Methods Employed:

The data processing requirements of the Gerontology Research Center have been analyzed in depth, and a system is being specified through utilization of standard engineering techniques.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

None

Proposed Course:

The computer system, once specified, will be procured on open bid. Limited system software generation will be necessary to provide for efficient use of the computer system.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.25
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Computer System for the Laboratory of Psychology, Section on Comparative Behavior, Poolesville

Previous Serial Number: None

Principal Investigator: Daniel Syed

Other Investigators: Martin Miller, Kenneth Kempner

Cooperating Units: None

Man Years:

Total: .1
Professional: .1
Other: 0

Project Description:

Objectives:

This project is designed to provide a real-time data acquisition computer system in support of on-line experiments being conducted by the Section on Comparative Behavior, Laboratory of Psychology, NIMH. Such a system is to provide control of sophisticated experimental equipment and also to permit extended periods of operation to be accomplished without operator intervention. Limited off-line computational capacity will be provided to meet requirements for statistical analysis.

Methods Employed:

A detailed study of the computational and control requirements of the Section on Comparative Behavior, NIMH, was performed, and standard engineering techniques are being used, in the design of a small computer system.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

None

Proposed Course of Project:

The computer system will be procured on open bid.

Honors and Awards: None

Publications: None

July 1, 1967, through June 30, 1968

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH

DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities 1. DCRT - 5

2. PHYSICAL SCIENCES LABORATORY 3. Dr. G. H. Weiss

I. SUMMARY

During the present reporting period, the Physical Sciences Laboratory grew from two professionals to ten. In consequence,not only has work gone forward at an accelerated pace on projects already started,but many new areas of investigations have been opened. These include the theoretical calculation of conformation of biologically interesting molecules, the development of statistical mechanical models of membrane transport phenomena, developments in the theory of nuclear magnetic resonance experiments, and the application of nonlinear numerical techniques for the resolution of data from biochemical separation experiments.

II. CURRENT BRANCH PROGRAMS

A. Objectives

To develop a group with theoretical interests to analyze biological phenomena in the terminology of theoretical physics and theoretical chemistry. To act as consultants in theoretical aspects of physics and chemistry to experimental scientists at NIH and to members of DCRT who may require such services in conjunction with their own work.

B. Progress of Current Programs

1. Development of Ultracentrifugation Theory

a. The program for generating accurate solutions to the Lamm equation which had been running on the 360/75 in New York has been transferred to the 360/75 in Greenbelt. Modifications have been made in it so that it now runs approximately twice as fast as in New York.

b. An extensive development of the theory and of relevant numerical solutions of the Lamm equation has been made for band centrifugation.

c. A theory has been developed for centrifugation in a density gradient, and relevant numerical solutions are being generated.

2. Theory of the Helix-Coil Transition in Polypeptides

a. Recent experimental developments indicate that the currently accepted theories of the helix-coil transition are incorrect, and do not describe the phenomenon even qualitatively. Further experiments on this phenomenon are being carried out, and a theory consistent with known experimental facts is being developed.

3. Computer Processing of NMR Data

a. Numerical procedures and programs have been developed for the interpretation of NMR data, particularly for the resolution of spectra. These programs are presently being used by members of NIAMD.

4. Molecular Model Building

a. Analysis of small molecules has been carried out, particularly on formic and acetic acids, and on the base-pairing problem, using potential functions developed during the last year. Excellent agreement has been obtained between present calculations and thermodynamic and crystallographic data.

5. Biophysical Analysis

a. A model has been developed for the effects of noise on nerves described by equations similar to those of Hodgkin and Huxley. This model predicts the shape of threshold firing curves as a function of the noise, as a function of bathing solutions, and certain drugs. Good correlation has been obtained between results of the theory and experimental results.

6. Excitation and Transport Properties of Fluids

a. A theoretical description of certain microscopic properties has been developed, and good agreement obtained with computer studies of the same problems at other laboratories.

7. Theory of Cell Membranes

a. The theory of electrostatics has been used to study the feasibility of suggested mechanisms for transport across membranes. It has been shown that ionic interaction with the low dielectric membrane is an important barrier to flow. However, it can be lowered considerably by highly polarizable material at points in the membrane.

8. Consulting Services

a. The analysis of experiments on the effects of varying concentration and time exposure of several antileukemic drugs on tissue cultures of human leukemic cells has been completed. The technique is a useful one, and provides results in a laboratory setting consistent with certain clinical

findings. Hence this type of experiment can be used for a study of other possible parameters in the use of antileukemic drugs.

b. Further investigations on nonlinear curve fitting techniques are being undertaken, in connection with the continuing input of biochemical data. In particular, Dr. James Ferretti is collaborating with several members of NIAMD in the preliminary specification of programs and options for a computer to be used exclusively for processing data from the laboratories in Building 2.

9. Fundamental Studies

a. A theory of the kinetics of red cells in transient conditions such as after acute hemolysis, has been developed. The theory is consistent with known experimental facts and will be useful for the interpretation of labelling experiments after it is further checked out.

b. Accurate calculations of the diffusion constant of rod-like polymer molecules have been made. It has been shown that the commonly accepted Kirkwood-Riseman theory leads to spurious results in the region of physical interest.

c. The stochastic theory of the kinetics of chemical reactions has been shown to lead to the same result as the deterministic, or mass-action theory, in the thermodynamic limit.

d. Preliminary calculations have been made for nuclear magnetic resonance in the presence of two radio frequency fields.

e. It has been shown that over 100 leads are required for an accurate measurement of the equivalent heart dipole.

C. Program Plans

1. Development of Ultracentrifugation Theory

Work will be begun on numerical solution of the Lamm equation for associating systems, and for density gradient centrifugation. Hopefully, work will be completed on the analysis of centrifugation in the presence of large pressure gradients.

2. Molecular Model Building

Theoretical conformations will be calculated for crystals and a start will be made on similar calculations for several protein molecules.

3. Biophysical Analysis

The work done to date suggests that a more general theory of noise in nonlinear systems is required for the understanding of much interesting nerve behavior. Hence similar studies will be undertaken on other model systems.

4. Theory of Cell Membranes

Calculations will be undertaken to ascertain the effects of electrostriction and carrier molecules on membrane transport.

Serial No. 5.1
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 20, 1968

Project Title: Theory of the Ultracentrifuge

Previous Serial Number: 5.1

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: Ralph Nossal, Ph.D., Richard I. Shrager, David Yphantis, Ph.D.

Cooperating Units: Charles W. Boone, M.D., NCI

Man Years

Total:	.70
Professional:	.50
Other:	.20

Project Description:

Objectives:

To determine the effects of various factors such as concentration dependent sedimentation, pressure, density gradients, variations in rotor speed, and polydispersity on current techniques for determining molecular weights. To devise corrections and new techniques of ultracentrifugation which bypass or eliminate these effects.

Methods:

The methods employed include numerical solutions to partial differential equations and classical analysis.

Major Findings:

Formulae have been developed for the time to equilibrium in density gradient centrifugation. Preliminary calculation of optimal density gradients and other relevant parameters to separate tumor cells in a Ficoll gradient have been successful. The theory of band centrifugation in non-ideal solutions has been developed.

Significance to Biomedical Research

Density gradient centrifugation is one of the most significant modern tools for separating molecular species and for populations of macroscopic tumor cells. Band centrifugation is of considerable importance in the study of biologically interesting molecules since very little material is required for such experiments.

Honors and Awards: None

Publications:

Weiss, George H. (with I. Billick, M. Schulz): Quasi-equilibrium Experiments with Rotor Deceleration, Journal of Physical Chemistry, 71: 2496-2502, 1967.

Weiss, George H. (with M. Dishon, D. Yphantis): Numerical Solutions to the Lamm Equation III. Velocity Centrifugation, Biopolymers, 5: 691-713, 1967.

Weiss, George H. (with I. Billick, M. Dishon, D. Yphantis): Numerical Solutions to the Lamm Equation IV; Rotor Slowing Experiments, Biopolymers 5: 1021-1028, 1968.

Weiss, George H. (with M. Dishon, D. Yphantis): Numerical Solutions to the Lamm Equation V: Band Centrifugation, Proceedings of the New York Academy of Sciences Conference on Ultracentrifugal Analysis (to appear).

Serial No. 5.2
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Theory of the Helix-Coil Transformation in Polypeptides in Solution

Previous Serial No.: None

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Livio Paolillo, Ph.D. (NIAMD)

Cooperating Units: NHI, NIAMD

Man Years:

Total:	.45
Professional:	.40
Other:	.05

Project Description:

Objectives:

To determine the effects of various factors such as Van der Waals interactions, hydrogen bonding, and electrostatic effects on the stability of helical polypeptides. To formulate models for the helix-coil transformation by the application of statistical thermodynamics.

Methods:

Applying NMR spectroscopy as the experimental approach towards evaluating the relative importance of the various interactions and then comparing the theoretical calculations with experimental data.

Major Findings:

It is possible to quantitatively determine the helix-coil concentration ratio under various experimental conditions and to completely follow the helix-coil transformation in various polypeptides. These results clearly showed that previous models do not adequately describe the phenomenon.

Significance to Biomedical Research

Polypeptides serve as protein model compounds and knowledge of the helix-coil transformation is essential to a complete understanding of the phenomenon of protein denaturation.

Honors and Awards: None

Publications:

Ferretti, James, A.: 100 Mc./sec Nuclear Magnetic Resonance Study of the Helix-coil Transformation in Polypeptides, Chemical Communications, 1030 (1967).

Serial No. 5.3
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Computer Processing of Nuclear Magnetic Resonance (NMR) Spectral Data

Previous Serial Number: None

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Mildred McNeel

Cooperating Units: None

Man Years

Total:	.45
Professional:	.40
Other:	.05

Project Description:

Objectives:

To develop methods, using the quantum mechanical NMR spin Hamiltonian, to analyze complex NMR spectra. To devise methods for computer simulation of experimental spectra and to decompose experimental spectra into their individual component frequencies and intensities.

Methods:

Iterative computer techniques are employed to approximate experimental NMR spectra by least squares fits.

Major Findings:

It is possible to completely analyze complex NMR spectra with the use of a computer. These results provide means for testing the sensitivity of the experimental spectrum changes in the various parameters.

Significance to Biomedical Research:

NMR spectra provide fundamental information about the structure and conformations of biologically interesting molecules.

Honors and Awards: None

Publications:

Ferretti, James A. (with E. Lustig, E. P. Ragelis, and N. Ivy), The Analysis of AA'BB' Nuclear Magnetic Resonance Spectra by Weak Double Irradiation. Application to Two Isomeric Cyclobutane Derivatives. Journal of the American Chemical Society 89, 3953 (1967).

Serial No. 5.4
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Molecular Model Building Using Theoretically and Empirically Determined Intra- and Intermolecular Potential Functions

Previous Serial Number: 3.14

Principal Investigator: William P. Minicozzi, M.D.

Other Investigators: Dan Bradley, Ph.D. (NIMH), James Kiefer, Ivan Darvey, Ph.D. Virginia Aandahl, Howard Nash, M.D. (NIMH).

Cooperating Units: Laboratory of Neurochemistry, Section on Physical Chemistry, NIMH

Man Years:

Total:	1.0
Professional:	1.0
Other:	0.0

Project Description:

Objectives:

Elucidation of the nature of protein folding.

Methods Employed:

X-ray crystallographic and electron scattering data are used to determine bond lengths and fixed bond angles for the molecular system under consideration. Then a linear combination of atomic orbitals is used to calculate charge distributions in the molecule under consideration among the various atoms consistent with the experimentally determined dipole moment of the molecule. An atom-atom interaction energy is the sum of four components in our model:

1. Electronic repulsion due to overlap of orbitals.
2. Electron correlation energy arising from the induced instantaneous dipoles.
3. Electrostatic energy, calculated by making a monopole-monopole approximation. Values of atomic monopoles are obtained as described above.

4. Static induced dipole energies calculated from an approximate formulation, which becomes more exact as the internuclear distance increases. Using quantum mechanical and classical mechanical derivations for functions and empirically determined values for some of the derived parameters, atom-atom potential functions are constructed. We then use these potential functions to calculate atom-atom interaction energies as a function of interatomic distance. Finally, the molecular system under consideration is allowed to undergo all possible internal rotations and assume a variety of intermolecular configurations. The energy of each is calculated by summing overall atom-atom interactions. In many instances, the accuracy of these procedures may be checked by comparing calculated results with experimentally determined energies and configurations.

The calculations necessary for these procedures require the development and testing of linked computer programs.

We are in the early stages of displaying molecules on an oscilloscopic screen. This will be an important part of the total procedure.

Major Findings:

Approximate intermolecular potential functions can be used quite effectively to predict the energies and conformations of simple molecular systems. They have not yet been tried on large molecular systems (e.g. proteins).

Significance to Biomedical Research:

Accurate prediction of the energies of molecular configurations enable one to know whether specific chemical reactions will or will not occur and why. This information should be of considerable importance to chemotherapy.

Honors and Awards: None

Publications: None

Serial No. 5.5
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Biophysical Analysis

Previous Serial Number: None

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: None

Cooperating Units: Harold Lecar, Ph.D., Biophysics Laboratory, NINDB.

Man Years:

Total:	.40
Professional:	.35
Other:	.05

Project Description:

Objectives:

To relate the fluctuations in firing thresholds of nerves to the chemical and physical processes underlying excitation. To analyze existing data, in order to test theories concerning mechanisms of transport of ions across nerve membranes.

Methods:

Existing neurophysiological equations (the Hodgkin-Huxley equations) have been modified to include fluctuating forces. Methods of non-linear mechanics and statistical physics have been applied to analyze the equations.

Major Findings:

It has been possible to relate the functional shape of the threshold firing curve, and the width of the threshold region, to models of thermal noise and conductivity fluctuations. There is also good correlation between the theory and results of certain experiments in which temperature, bathing solutions, and drugs have been used to change the threshold properties of nerve axons.

Significance to Biomedical Research:

It is hoped that the study will facilitate better understanding of the physical processes underlying the excitation and propagation of nervous impulse.

Publications: None

Serial No. 5.6
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Physico-chemical Properties of Macrocyclic Antibiotics

Previous Serial Number: None

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: James A. Ferretti, Ph.D.

Cooperating Units: Gerald Ehrenstein, Ph.D., Biophysics Laboratory, NINDB

Man Years:

Total:	.25
Professional:	.25
Other:	.00

Project Description:

Objectives:

Various macrocyclic antibiotics (e.g. Gramicidin, Valinomycin, Nonactin) facilitate transport of monovalent cations across biological structures such as mitochondrial and plasma membranes. This project is directed towards understanding the molecular mechanisms underlying facilitated transport and, in particular, the ion selective aspects of the latter.

Methods:

Physical probes such as analytic ultracentrifugation, high resolution NMR, and laser light scattering are being applied to determine the molecular structure and properties of these antibiotics. When necessary, existing physical theory is extended in order to support suitable modifications of experimental technique. Reconstituted phospholipid bilayers are used in conjunction with these studies, and new experimental procedures are developed when appropriate.

Major Findings:

The project has only recently been undertaken; no major results have yet been obtained.

Significance to Biomedical Research:

It is hoped that these studies will lead to better understanding of the details of the pharmacological action of macrocyclic antibiotics. Further, the latter may be analogs of cation transport "channels" in excitable tissues, and it is of some interest to establish their relevancy in this regard.

Publications: None

Serial No. 5.7
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Excitation and Transport Properties of Fluids

Previous Serial Number: 5.4

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years

Total:	.45
Professional:	.40
Other:	.05

Project Description:

Objectives:

To provide basic knowledge concerning the excitation properties and transport properties of fluids, to provide theoretical explanations of experimental evidence of high frequency collective atomic motions of simple classical fluids.

Methods:

Theoretical techniques of mathematical physics and statistical mechanics are employed, in order to develop new physical theories.

Major Findings:

A theoretical description of the properties of fluids has been developed and its consequences are being explored. Preliminary analysis indicates good agreement with results of neutron scattering experiments on simple fluids.

Significance to Biomedical Research:

Almost all biological phenomena occur in a fluid environment. A number of fundamental questions concerning the physical behavior of fluids yet remain unanswered. Their elucidation will ultimately facilitate better understanding of the functions and properties of biological systems.

Publications:

Nossal, R. (with R. Zwanzig), Approximate Eigenfunctions of the Liouville Operator in Classical Many-Body Systems. II. Hydrodynamic Variables, Physical Review 157, 120 (1967).

Nossal, R.: Collective Motion in Simple Classical Fluids, Physical Review 166, 81 (1968).

Nossal, R.: Hamilton's Principle for Continuum Viscoelestic Fluids, J. Mathematical Physics (to appear).

Serial No. 5.8
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Intermolecular Forces Acting in the Cell Membrane

Previous Serial No.: None

Principal Investigator: V. Adrian Parsegian, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years:

Total:	.55
Professional:	.5
Other:	.05

Project Description:

Objectives:

To identify and calculate those intermolecular forces important in stabilizing the cell membrane and in determining transport of materials across the cell boundary. So far these have been coulombic interactions between charged species, short-range interactions of the van der Waals type, and charge interactions with media of different shapes and dielectric properties.

Methods:

Classical and quantum-mechanical treatment of electromagnetic and statistical-mechanical behavior as well as molecular models of specific molecular associations

Major Findings:

Van der Waals forces act as an interfacial tension between lipid aggregates and aqueous medium. Ionic interaction with the low dielectric membrane is an important barrier to ion flow; it can be lowered by the presence of highly polarizable material at points in the membrane.

Significance to Biomedical Research:

The transport of ionic and neutral species across the cell membrane is strongly dependent on interaction with the membrane material as well as on the structure of that material.

Honors and Awards: None

Publications:

Parsegian, V. A.: An Energetic Model of Ionic Lipids in the Liquid-Crystal State, Proceedings of the International Conference on Biological Membranes, Frascati, Italy (to appear).

Serial No. 5.9
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 20, 1968

Project Title: Consulting Services

Previous Serial Number: 5.3

Principal Investigators: George H. Weiss, Ph.D., Richard I. Shrager

Other Investigators: Mildred McNeel, Yashar Hirshaut, M.D., Todd Miles, Ph.D., Frank Howard, Ph.D., Donald Young, M.D.

Cooperating Units: Medicine Branch, NCI, Laboratory of Molecular Biology, NIAMD, Clinical Center

Man Years:

Total:	1.1
Professional:	1.0
Other	.1

Project Description:

To provide consulting services in biometry, applied mathematics, theoretical physics and chemistry to workers who are primarily in experimental fields.

Methods:

The methods include statistical analysis, theoretical physics and chemistry, and applied mathematics.

Major Findings:

The use of human leukemic cells in culture has been shown to be useful in describing the effects of antileukemic drugs *in vivo*. Experiments indicating the effects of varying concentration and time of contact of drugs with these cells indicate that tissue cultures can be used as an assay device, and to suggest new drug regimes in leukemia.

Numerical techniques have been devised to correlate data from circular dichroism and from optical rotatory dispersion experiments, and to decompose them into spectral components.

Methods:

Standard statistical techniques have been used for the first project. Methods for nonlinear curve fitting and for the evaluation of Kramers-Kronig relations have been devised for the second project.

Honors and Awards: None

Publications: None

Serial No. 5.10
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Fundamental Studies

Previous Serial Number: 5.4

Principal Investigators: George H. Weiss, Ph.D., James A. Ferretti, Ph.D., Ralph Nossal, Ph.D., Gershom Zajicek, M.D.

Other Investigators: Livio Paolillo, Ph.D., Irwin Oppenheim, Ph.D. and Robert Zwanzig, Ph.D.

Man Years:

Total:	4.15
Professional:	3.7
Other:	.45

Project Description:

This project encompasses several lines of investigation, some, but not all of which relate to biomedical problems. These include theoretical studies of red cell kinetics, the calculation of diffusion properties of long chain polymers, the comparison between deterministic and stochastic theories of chemical reactions, development of the theory of nuclear magnetic resonance, design studies for a new electrocardiograph, and basic studies in statistical mechanics.

Major Findings:

A theory has been developed for red cell kinetics after acute hemolysis. The theory is phrased in terms of measureable parameters like red cell volume and osmotic fragility, and is in accord with presently known facts about red cell kinetics. Further experiments are needed to develop some of the necessary parameters of the model.

It has been shown that the classical calculation of the diffusional properties of rod polymers is incorrect, and that a more accurate account of boundary conditions is required in the hydro-dynamic theory.

For some time it has been conjectured that the stochastic theory of chemical kinetics would predict the existence of macroscopic fluctuations not due to experimental error. This has now been settled, as it has been shown that the fluctuations from results predicted by the deterministic theory are negligible.

A density matrix formulation for NMR relaxation in the presence of a second radio frequency field has been shown to be feasible.

Calculations of the number of leads required for accurate determination of the equivalent heart dipole indicate that a number of the order of 150 is necessary to achieve errors of less than 15%.

Models have been developed for the decay of correlations in simple mechanical systems with large numbers of degrees of freedom. It has been shown that the introduction of coupling does not change any of the qualitative results of the theory developed for independent particles.

Methods:

The methods used include partial differential equations, statistical theory, classical analysis, numerical analysis, and other techniques commonly used in theoretical physics.

Significance to Biomedical Research:

The work in red cell kinetics may lead to new interpretations of labelling experiments. The work in polymer diffusion is constantly used in the interpretation of experiments on DNA. A clear delineation of the region of validity of the theory will be of use in the interpretation of discrepancies between experiments and theory. The work on the electrocardiograph may lead to one which will be more useful for pediatric cardiology.

Awards and Honors: None

Publications:

Ferretti, James A. (with L. Paolillo): The Signs of Geminal and Directly Bonded Hydrogen-Phosporus Coupling Constants. Ricerca Scientifica CCSS-11, 103-105, 1967.

Weiss, George H. (with H. Pettigrew): Epidemics with Carriers, The Large Population Approximation. J. Appl. Prob. 4, No. 2, 257-286, 1967.

Weiss, George H. (with J. Gart): Graphically Oriented Tests for Host Variability in Dilution Experiments. Biometrics 23, 269-284, 1967.

Weiss, George H. (with I. Oppenheim and K. Shuler): On the Decay of Initial Correlations in Stochastic Processes. J. Chem. Phys. 46, 4100-4114, 1967.

Weiss, George H. (with I. Billick and M. Schulz): Quasi-equilibrium Experiments with Rotor Deceleration. J. Phys. Chem. 71, 2496-2502, 1967.

Weiss, George H.: On the Master Equation Formulation of Chromatography Theory. Separation Science 2, 551-554, 1967.

Weiss, G. H.: An Introduction to Lattice Dynamics. In Many-Particle Physics. ed. by E. Meeron (Gordon & Breech, 1967) 129-182.

Weiss, G. H.: First Passage Time Problems in Chemical Physics in Advances in Chemical Physics XIII. (Wiley & Sons, New York) 1-18, 1967.

Weiss, George H.: (with I. Oppenheim and K. E. Shuler): Stochastic Theory of Multistate Relaxation Processes. Adv. in Molecular Relaxation Proc. 1, 13-68, 1967

Weiss, George H. (with R. Hoye, J. Gart, Ketcham, G.) Potentiation of Laser Oncolysis with Pretreatment X-irradiation. Rad. Res. 32, 112-117, 1967.

Weiss, George H.: The Intersection Delay Problem with Gap Acceptance Function Depending on Space and Time. Trans. Res. 1, 367-371, (1967).

Weiss, George H. (with Dishon, M. and Yphantis, D. A.) Numerical Solutions to the Lamm Equation III. Velocity Centrifugation. Biopolymers 5, 691-713, 1967.

Weiss, George H. (with Billick, I., Dishon, M., and Yphantis, D. A): Numerical Solution of the Lamm Equation IV; Rotor Slowing Experiments. Biopolymers 5, 1021-1028, 1967.

Weiss, George H. (with Nossal, R.): Analysis of a Generalized Ladenburg-Reiche Integral. J. Quantit. Spect. 8, 763-771, 1968.

Weiss, George H.: Equations for the Age Structure of a Growing Population, Bull. Math. Biophys. (to appear).

Weiss, George H.: An Introduction to the Statistical Theory of Irreversible Processes, In Irreversible Thermodynamics (to appear).

Weiss, George H. (with R. Zwanzig, J. Kiefer): On the Limits of Validity of the Kirkwood-Riseman Theory, Proc. Nat. Acad. Sci. (to appear).

Weiss, George H. (with Dishon, M. and Yphantis, D. A.) Numerical Solutions for the Lamm Equation. V. Band Centrifugation. Proc. Conf. on Ultracentrifugal Analysis (to appear).

Weiss, George H. (with Grover, N.): Theoretical Distribution of Two-Dimensional Projections in Electronmicroscopic Sections. Proc. Nat. Acad. Sci. (to appear).

Weiss, George H.: Some Models for the Decay of Initial Correlations in Dynamical Systems, Adv. in Chem. Phys. (to appear)

Zajicek, Gershom: A Computer Model Simulating the Behavior of Adult Red Blood Cells, Journal of Theoretical Biology (to appear)

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT - 6
Serial Number

2. HEURISTICS LABORATORY

3. James R. Slagle
Chief

I. SUMMARY

This Laboratory which was activated during this reporting period is currently pursuing two broad objectives which are intended to apply the techniques from computer research to problems in biomedical research. The projects which embody these objectives are qualitatively described. The two projects include the development of programs for answering questions and the design and development of a computing system for the processing of biological images. Both of these projects are still in the initial stages of development.

II. CURRENT LABORATORY PROGRAMS

The Heuristics Laboratory which was activated half way through this reporting period has been formed from newly recruited personnel and the Biological Image Processing group which was transferred from the Computer Systems Laboratory. This Laboratory is presently composed of 12 scientific and 2 clerical individuals. The Laboratory will use as its main computing system the PDP-10 computer which was purchased during the reporting period.

A. Objectives

This Laboratory is currently pursuing two broad objectives:

1. To develop concepts, techniques and programs for the automatic answering of questions and the solving of problems.

2. To design and develop a computing system and the associated programming techniques for the on-line, interactive processing of biological images.

B. Progress

The objectives previously set forth represent the current broad goals of the Laboratory. Within each broad objective individuals have begun projects which realize relatively separable sub-goals. Greater detail is supplied in the individual project reports.

Progress Toward Objective #1 (To develop concepts, techniques and programs for the automatic answering of questions and the solving of problems.)

There are many problem areas within the biomedical environment in which facts emerge and deductive reasoning from the facts must be done. Various studies in the past have shown that the computer can be used to do deductive reasoning. Two serious difficulties prevent the application of computers to realistic problems. They are:

1. The formulation of a question has a very important influence on the answerability. A poor formulation can make a question effectively unanswerable.

2. As the number of facts in the data base increases, the time required for doing deductive reasoning increases beyond practical limits.

In order to overcome the first difficulty, studies in the formulation of questions and problems have begun. The outlook for these studies is a developing theory of question and problem types. As distinct types of questions and problems arise, computer programs are being written. Work towards solving the second difficulty has been concerned with developing strategies for determining the relationship of facts to one another and to the process of utilizing the facts. The intention is to use only relevant facts in the answering of a question and the solving of a problem. Greater detail is provided in Individual Project Report DCRT 6.1.

Progress Toward Objective #2 (To design and develop a computing system and the associated programming techniques for the on-line, interactive processing of biological images.)

Work has been completed on the negotiation of a contract with the Digital Equipment Corporation for the purchase of the Laboratory's PDP-10/50 computer-display system. The negotiations have been completed and the computer is scheduled for delivery in November 1968. The software required for providing an on-line, reactive system for biological image processing is being designed. It has been decided to use the list processing language LISP as the high level language for integrating and coordinating the various components of the system. When the system becomes operational, we expect that biologists and other scientists with image processing problems will use it in collaboration with computer specialists of this Division to develop specific, problem-oriented algorithms. Greater detail is provided in Individual Project Report DCRT 6.2.

General Remarks

The staffing of the projects in this Laboratory has been hindered by the difficulties in obtaining adequately experienced personnel. As the role of the Laboratory expands through the acquisition of the PDP-10 computer and the initiation of other projects, further recruiting will have to occur. While space for the computer has been found, the location of the space leaves much to be desired. The biological image processing computer facility should be spatially proximate to the research activities. The role of this computer is not a peripheral data processor but an integral part of the scientific process.

The space for the present staffing of this Laboratory is in part inadequate. As additional individuals are added to this Laboratory, it becomes imperative

to find additional space. An effort should be made during the next reporting period to integrate the computing facility with an enlarged block of office space.

Serial No. DCRT - 6.1
1. Heuristics Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Automatic Question-Answering

Previous Serial Number: None

Principal Investigator: James R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D., John K. Dixon, Richard J. Feldmann, Deena A. Koniver, Richard C. T. Lee, Ph.D.

Cooperating Units: None

Man Years
Total: 5.1
Professional: 3.8
Other: 1.0

Project Description:

Objectives:

1. To develop concepts and formalisms for the representation of facts, questions and problems.

2. To develop strategies which reduce the amount of computation required to answer questions and solve problems.

Methods Employed:

As new concepts, formalisms and strategies arise, internal papers are generated. The computational feasibility of these proposals are studied and computer programs are written and debugged. Presently, the Q-32 computer in Santa Monica, California is being used. The computer language LISP which is a laboratory standard is also being utilized on the IBM 360/50 computer.

Major Findings:

A computer program written to test a new inference rule for facts in question-answering has shown that significantly fewer working facts are

generated during the course of the answering of a question. Saturation of computer memory in the past has been one of the factors which prevented the development of a practical question-answering system. A hypothesis concerning the generation of models for the mechanization of inductive inference was tested by a computer program. The results indicated that too many instances of a model could occur. A reformalization and testing of the General Problem Solver concepts has indicated that certain problems which are effectively unanswerable with the inference rule type programs can be answered. The concepts used to decide effective courses of action in a multipurpose search program have been applied to the environment of question-answering by resolution.

Significance to Biomedical Research and the Program of DCRT:

Several demonstrations of a question-answering program written during this reporting period have been made to members of other groups in DCRT. A fully developed question-answering-problem solving program would allow the working scientist to turn over to the computer routine questions and problems. The computer would solve these routine matters while the scientist was attending to the more important and difficult problems.

Proposed Course of Project:

1. The development of new methods for question-answering will continue. Whenever possible, programs will be written to test these ideas.

2. A collaborative project with the Division of Research Facilities and Resources will most probably begin. In this project deductive question-answering techniques will be applied to pharmacological data embedded in qualitative models of human anatomy and physiology.

3. As the PDP-10 computer becomes available programs and activity will be transferred from the existing computer services.

Honors and Awards: None

Publications:

Slagle, James R.: Automatic Theorem Proving with Renamable and Semantic Resolution, Journal of the ACM, Vol. 14, 4, (Oct. 1967), 687-697.

Slagle, James R.: Generalizations of a Complex Analogue of the Real Tchebichev Polynomial Theorem, American Mathematical Monthly, Vol. 75, 1, (Jan. 1968), 58-59.

Slagle, James R.(with P. Bursky): Experiments with a Multipurpose, Theorem-Proving Heuristic Program, Journal of the ACM, Vol. 15, 1, (Jan. 1968), 85-99.

Chang, C. L.: Fuzzy Topological Spaces, to appear in Journal of Mathematical Analysis and Applications.

Serial No. DCRT 6.2
1. Heuristics Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1967 through June 30, 1968

Project Title: Biological Image Processing

Previous Serial Number: DCRT - 4.7

Principal Investigators: H.W.Vreenegoor, DCRT
Lewis E. Lipkin, M.D., NINDB

Other Investigators: Malcolm C. Bruce, Sam Bryan,
Edward C. Hill, Louis Hodes, Ph.D.

Cooperating Units: Section on Pathology, PRB, NINDB

Man Years
Total: 6.3
Professional: 6
Other: 0.3

Project Description:

Objectives:

1. Development and utilization of new techniques and devices to permit the study of biological images in order to determine significant characteristics of these images for classification and identification purposes.

2. Research directed toward the development of a "natural language" communication system to optimize biologist-computer interactions in the context of image processing.

3. Study, evaluation, and implementation of techniques permitting semi-automatic and/or automatic processing of large volumes of biological images.

Methods Employed:

A number of computer programs were developed on the IBM 360 computer for transforming digitized picture data to picture data compatible with the Stromberg-Carlson 4020 microfilm printer.

A picture processing oriented language, called PAX, was developed

under contract by the University of Maryland for the IBM 360 computer.

An on-line reactive approach, using LISP as the high level language, was studied to determine its feasibility and utility for picture processing. Members of this group have worked on the negotiation of a contract with the Digital Equipment Corporation for the purchase of a PDP-10/50 computer-display for use by the laboratory. The contract has been signed. The system is scheduled for delivery in November 1968.

Major Findings:

Experience with the Stromberg-Carlson 4020 has shown that "natural" looking biological pictures can be reproduced from digitized data. The process is most useful for producing occasional high quality prints or film. It is too cumbersome and costly for large numbers of pictures.

The PAX picture processing language has provided a capability for analyzing pictures in the batch processing mode. Its primary function will be the determination of the usefulness of newly developed picture processing algorithms.

Studies of the feasibility of using on-line, interactive LISP with PAX-like processes, show it to be an important step towards achieving a system for rapid on-line development of picture processing algorithms.

Significance to Biomedical Research and the Program of DCRT:

The project provides resources required for the study of biological images in an objective and repeatable manner. The interaction of the biologist with selected and displayed images should provide information important towards acquiring a better understanding of the underlying structure. The project applies state-of-the-art computing devices and techniques to an important segment of biological investigation.

Proposed Course of Project:

1. Development and implementation of a LISP based on-line reactive system for the PDP-10/50 with the following major characteristics:

    a. capability of creating and handling display lists using the modified DEC 340 display

    b. capability to handle PAX or PAX-like picture processing techniques

    c. capability to handle FORTRAN and assembly language generated programs

    d. capability to accept Rand Tablet data

2. Development of specific picture processing algorithms.

3. Study of the utility of a flexible, programmable flying spot scanner for the project.

Honors and Awards: None

Publications: None

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY
Report of Program Activities
July 1, 1968, through June 30, 1969

ANNUAL REPORT

TABLE OF CONTENTS

Page

July 1, 1968 thru June 30, 1969

PUBLIC HEALTH SERVICE-NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

1. DCRT-2
Serial Number

2. COMPUTER CENTER BRANCH

3. J. D. Naughton
Branch Chief

I. SUMMARY

During the previous two years (FY '67 & '68) the major activity of the Computer Center Branch (formerly part of the Computation and Data Processing Branch) had been the conversion of all NIH computer activity from second generation to new third generation computer equipment. As the task itself was completed it became obvious that if the NIH biomedical research and management programs were to recognize the maximum benefit from the recent developments in computer science technology a completely new philosophy and set of goals would have to be established. To accomplish this the Computer Center Branch was established with the general responsibility to develop a powerful central computer system and provide the necessary standards and training to NIH scientific and administrative personnel so that this tool could be most effectively applied to the NIH mission.

In August the Computer Center became the first IBM 360 installation in the world to implement a computer system where three individual computers in the same room all had simultaneous access to a common disk unit on which was stored systems software and application programs.

This innovation allows any of three computers to process any job, reduces the need for disk storage, and also improved the thru-put of the system approximately 15% at no additional cost. Further efficiencies were introduced in March when a multi-programming system (a multiple number of jobs running in a computer at the same time) was installed. The NIH central computer system now contains 4,000 on-line catalogued programs in six languages; operates 24 hours a day supporting over 600 scientific and data processing users running approximately 750 jobs each day. An interactive graphic display facility (television tube) was added to the central system to allow the scientist to visually inspect and manipulate biomedical data, such as molecular structures, where visual appreciation is vital. Progress was made in biological image processing thru the study of man-machine interaction with displays and the implementation of syntax-directed, picture-processing techniques which lead to a method of successfully representing relationships and directing the processing of picture thru the use of grammars. Significant progress was made during the year with the installation of seven remote job entry terminals (IBM 360 model 20's) in various Institutes and Divisions. These RJE terminals are connected to the NIH central computer system via standard telephone lines and provide the

researcher the full power of the central system, but without the penalty associated with leaving his own research environment. Even though the RJE terminals have been used for less than a year, over 200 jobs a day are processed through these RJE terminals. During the latter part of the year a conversational programming system (CPS) was introduced which allows multiple users to simultaneously write, debug, and execute computer programs from typewriter-like terminals installed throughout the campus and connected to the central system via standard telephone lines. Although the system is still quite new and only 15 terminals have been installed, some 75 jobs are received daily from CPS terminals.

In order to maintain rapid service and insure efficient use of the system, programming standards have been developed and published as a "Users Guide to the Computer Center." An "Operators Guide" insures that all jobs are handled properly at the Center. Both of these documents have been adapted for use in other federal agencies and in the private sector. In order to make the facilities of the Center more widely understood a Computer Training Program with a curriculum of 28 courses was developed and taught twice this year to over 800 students. An additional improvement in communications between the Center and its many users is represented by a monthly newsletter, INTERFACE, published by the Center to describe overall plans, accomplishments, services and policies which now enjoys a world-wide circulation of over 1250.

II. BRANCH PROGRAMS

A. Objectives

The newly established Computer Center Branch assumed the responsibility of planning, designing, implementing and maintaining all hardware and software systems necessary for NIH's central computer facility. The overriding purpose is to provide a computer facility to most effectively meet the dynamic and diverse requirements of the NIH research investigators and management. The most powerful research tool in generations is computer processing. It is the Computer Center Branch's objective to put this tool into the hands and office of every researcher at NIH. This will be accomplished by providing a strong central computer complex, with sufficient capacity and software to service the NIH needs. The central complex will be surrounded by a variety of terminals, ranging from typewriter-like terminals to small computers, all communicating with the central computers over phone lines. This enables the full power of a large computer, once isolated in a single place, to be brought directly into the researcher's office or laboratory. To further this end, the Computer Center Branch will define and provide training courses for NIH personnel in computing concepts and in the use of the central computer system. Standards have been, and will continue to be, developed to encourage a high level of efficient and effective use of the central computer system, as well as to guide and assist all NIH users of the central computing facility. All facets of data processing will be made available to all Institutes and Divisions of the NIH. Research efforts will continue to develop new computer system techniques and languages which will enable the Computer Center Branch to better meet the specific

requirements of the NIH environment. Technical knowledge and systems experience will be exchanged with other organizations engaged in computation in support of biomedical research and administration to keep the Computer Center Branch at the forefront of biomedical computing.

B. Current Program Progress and Accomplishments

During the past year two areas were attacked extensively while research into other promising areas continued. Meeting the seemingly insatiable demands of Computer Center users for capacity and rapid turnaround time was worked on most extensively, culminating in the introduction of two forms of remotes and the installation of the MVT (Multiprogramming with a Variable Number of Tasks) version of OS/360. These systems improvements markedly reduced turnaround time and improved thruput, meaning faster service for users. We succeeded in bringing the computer into the laboratory with the installation of remote terminals enabling both a conversational system (CPS) and the Remote Job Entry System. The MVT system allowed the same amount of computer power to be simultaneously applied to several jobs, thus increasing overall efficiency. The second area to receive major attention was the closing of the communications gap between the Computer Center and its users. The formation of the PAL Unit (Programmer Assistance and Liaison) gave the users the personal assistance needed if they were to make effective use of the facilities. The publication of INTERFACE kept the users informed of overall plans, new services and facilities, and abreast of current technical information. The issuing of the "Users Guide to the Computer Center" gave the user a complete picture of computing at NIH and its standards, while the extensive training courses that were established helped teach the user how to obtain the greatest return on their investment in computer processing.

Research and development work into extensible languages and image processing continues along with the development of new terminal systems that will put the computer into more labs and offices. Graphic display systems and their application to biomedical research are also being thoroughly investigated.

The following are examples of programs installed and underway.

1. Technical Newsletter, INTERFACE

During the year a project was established to organize and publish a periodic newsletter dedicated to improving communication and fostering collaboration among the NIH computer users. The Computer Center relies heavily on INTERFACE to disseminate current technical information and to describe overall plans, new services and facilities, and policies of general interest to the widening community of computer users. Ten issues have been published this year and were distributed, by request, to all computer users at NIH, most federal departments and many countries of the world. If it were not for the delays and difficulties encountered in the actual printing process itself, INTERFACE would have appeared more frequently.

2. Establishment of PAL Unit

As more complex computing facilities were made available to an increasing number of users with a wide variation of training and experience it became obvious that many users would require considerable personal assistance if they were to make effective use of these new facilities. In order to provide more responsive service to our users who were experiencing difficulty in using the computer, the Programmer Assistance and Liaison (PAL) Unit was established with the prime responsibility to provide users assistance in resolving problems encountered while using the system (software and hardware) maintained by the Center. This Unit has been exceptionally well received by NIH computer users and now provides some personal assistance to an average of 30 to 40 users every day.

3. New Computer Training Courses Offered

In the past, limited training courses in computer use have been available at NIH. The result was that many users called on only the most basic and obvious capabilities of the NIH computers when other or additional facilities might do more of their work for relatively less cost. This situation and the many advances since the original 360 installation made the need for a total EDP training program, oriented to the specific computing facilities available at NIH, absolutely essential if users are to obtain the greatest return on their investment.

The Computer Center took the first step toward meeting this critical need. Eighteen new courses have been designed to augment the knowledge of both scientific and business data processing programmers at all levels of experience, from novice to expert. Several courses are offered also for managers and computer operating personnel. Additional courses will be added as need is indicated and staff is available.

The brochure, "Computer Training Courses," presents detailed course descriptions, prerequisites, schedules and registration forms. As further steps toward closing this EDP training gap, DCRT labs and branches hope to increase offerings of specialty seminars in specific areas of computer applications--from mathematical modeling and curve fitting to query systems for data processing.

The courses were taught twice this year with a total attendance of over 800 students. In spite of the large number attending the courses it was still necessary to turn down over 500 NIH students due to lack of staff and classroom facilities. If these problems can be resolved we will increase the Computer Training Program next year.

4. Remote Job Entry Terminals (RJE)

Seven Remote Job Entry Terminals, IBM 360/20's, were installed in Institutes and Divisions of NIH during the year.

The RJE system allows any 360 job to be transmitted from the remote location directly to the central 360 system, to be processed along with other tasks, and the results transmitted back to the remote location for printing. The RJE user is provided the full power of the large central 360 system at his own location including all standard programming languages, libraries, cataloged procedures, and utilities normally available in the central system when 360 jobs are submitted at the Computer Center. Use of the RJE terminals requires no programming of Job Control Language changes.

The RJE terminal consists of three interconnected units; (1) a 600 card-per-minute card reader, (2) a 300 line-per-minute printer with 132 print positions using a 64-character print element and (3) a controller/transmission interface. The interface is connected to a "data set" (telephone company modulator/demodulator) which links the terminal through a standard telephone line (voice grade, 2400 baud) directly to the central 360 system in Building 12. Work is currently underway to increase the data transmission rates and thus allow even higher print and card read speeds at the terminal.

The operating schedule for the RJE terminals is 8:30 a.m. to 5:00 p.m. weekdays (excepting unscheduled downtime and preventive maintenance). Extended hours and weekend service has been requested but cannot be provided due to a severe shortage of operating personnel for the central system. Otherwise, the RJE service has been successful in all respects. There are an average of 260 computer jobs submitted daily via the seven existing terminals and three more are on order for installation next year.

5. Users Guide Published

A Users Guide to the services, standards, and use of the Computer Center was published and distributed in September. This Guide was prepared for the general NIH computing community and covers all facets of the use of the Computer Center, presenting the Center's services and equipment and the procedures for using them. The logistics of using the Center's 360 computers are described as well as the installation's standards pertaining to these services. For those programming for the 360 computers, it gives the program languages the Center supports along with the installation's programming standards used in preparation of 360 computer programs. A programmer's overview of the NIH Standard Operating System and a summary of Job Control Language (JCL) is included, as well as a large section on the Center's standard catalogued procedures and utility programs. This latter section on the Center's utility programs displays the most frequently used program decks for all supported languages. Lastly, computer hardware descriptions and the Center's equipment configurations are given.

This Users Guide is made available to all users of the Computer Center, to Institute and Division administrative or executive officers, to contract companies required to use the Computer Center and other organizations or individuals having a logical need for it.

The Computer Center has received over 100 written requests from other federal agencies and industrial concerns for copies of the Users Guide so

that it may be adapted, copied or used as an outline for preparing a similar document for use in their installation. Publishing the Users Guide, which forced the definition of standards, probably did more to improve the relationship, and satisfaction of the computer users than any other single activity of the Center.

6. Conversational Programming Introduced (CPS)

The Computer Center made available the first Conversational Programming System (CPS) on the NIH central system in December. After several months of testing and debugging the system was made available to NIH computer users in March. There are now 70 registered users of the CPS system and 15 keyboard terminals have been installed. As soon as difficulties in having telephone lines installed can be resolved another 20 terminals will be made available to users throughout NIH.

The CPS system allows multiple users to simultaneously write, debug and execute 360 computer programs conversationally. Programs are written in a subset of the PL/I language from an IBM 2741 selectric typewriter or a teletype terminal located in or near the user's office. The terminal communicates with the NIH central 360 system over telephone lines. This system is somewhat similar to the BASIC system currently available from commercial sources, but has the advantage that it will be less costly to use.

7. Pharmacology System

The system is being developed collaboratively by DRFR and DCRT as a research tool for pharmacology and toxicology. The project is a first step in DRFR's development and evaluation of an integrated set of computer techniques for handling data on chemical compounds and the functioning of living systems.

An Adage Graphics Terminal has been installed and the Computer Center is so structuring its effort that the computer systems developed can be generalized to support other areas of biomedical research and a wider community of users at NIH.

The terminal itself consists of a CRT display embedded in a fairly imposing hardware array. Basically, it displays three-dimensional picture information with great speed and directness. Beyond that, it is a system for the management of pictures, and for conversational interaction with the 360 system. Thus, a scientist with a three-dimensional computation problem, e.g., a crystallographer or pharmacologist, would use the terminal to observe the course of calculations, and approve, modify, or cancel them using the high-speed data connection with the 360. The terminal displays two-dimensional projections of the three-dimensional information which is stored in its memory, and the user is free to alter viewing angles, translations, and magnifications. The terminal is thus a general 3-D viewing device.

By fall, the programming system will be rounding out. This will include general purpose software in both the AGT and 360 to enable the user to manipulate his pictures and their underlying computations. Under the AGT-30 system, the picture appears as a solid object with a set of natural, easily manipulated viewing controls. The nature of operations will be very general.

Serial No. 2.1

1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: User Support and Communications

Previous Serial Number: None

Principal Investigator: F. Halverson

Other Investigators: Staff of PAL Unit

Cooperating Units: None

Man Years

Total:	7
Professional:	7
Others:	0

Project Description:

Objectives:

To provide the users of the Computer Center with the personal assistance necessary if they are to make effective use of the Center's facilities. To provide users assistance in resolving problems encountered while using the systems (hardware and software) maintained by the Center.

Methods Employed:

The PAL (Programmer Assistance and Liaison) Unit was established to perform this function.

Significance to Program of the Division:

The computer user at NIH has at his disposal a group of competent professional programmers that are able to assist him with his problems in running programs or using the computer facility. The PAL Unit notes all trouble areas and, through INTERFACE, communicates common problem areas to all users. In all facets of computer use the user has someone that can give him the answer or guarantee to find the answer for him.

Honors and Awards: None

Publications: None

Serial No. 2.2
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer Center Users Guide

Previous Serial Number: None

Principal Investigator: Robert H. Brunelle

Other Investigators: Staff of Systems Team and PAL Unit

Cooperating Units: None

Man Years

Total:	2
Professional:	2
Others:	0

Project Description:

Objectives:

To provide the users of the central computer facility with a guide to the services, standards and use of the Computer Center.

Methods Employed:

A 200-page Users Guide was published and distributed to all users of the Computer Center, to Institute and Division administrative or executive officers, to contract companies required to use the Computer Center and to other organizations and individuals having a logical need for it. Updates to the Users Guide are published and distributed as necessary to keep it current.

Significance to Program of the Division:

For the first time all information pertaining to computing was brought together in a single reference document. The Users Guide contains a description of all facilities and services and how to use them. Programming standards, languages supported, JCL summary and other facets of computing are all given in detail. It is a complete guide to computing for the computer user.

Honors and Awards: None

Publications: None

Serial No. 2.3
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Paper Tape Facility

Previous Serial Number: None

Principal Investigator: Robert H. Brunelle

Other Investigators: G. Lefelar, R. Hart

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

To provide a means by which investigators, whose laboratory gear produces paper tape, can have the information on these tapes processed by the central computing facility.

Methods Employed:

A paper tape reader and punch have been acquired and put on the 360 computer. A general-purpose program is being written that will transfer information from paper tape onto 9-track magnetic tape. The data can then be analyzed by 360 computer programs.

Significance to Program of the Division:

Laboratories will be able to acquire less-expensive gear on which to record their data, and still have the data analyzed by the central facility's computers.

Honors and Awards: None

Publications: None

Serial No. 2.4
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: HASP - Shared Spool Disks

Previous Serial Number: None

Principal Investigator: John Gilman

Other Investigators: None

Cooperating Units: None

Man Years

Total: 1/4
Professional: 1/4
Others: 0

Project Description:

Objectives:

1. To provide a single logical job input/output stream for all three IBM 360/50's (future 360/65) at DCRT-CCB.

2. To allow for more dynamic balancing of the work load among the multiple computers in the installation.

3. To provide an overall smoother work flow within the Operations Section of the branch.

4. To facilitate information gathering about the current status of work on the three systems.

5. To optimize the use of on-line storage devices for intermediate storage of job input and output.

Methods Employed:

The HASP spooling system currently in use on the 360/50's is to be modified to allow all three copies to use the same set of on-line spool disks.

Significance to Program of the Division:

The multiple computer environment at the Computer Center allows significantly better overall service to the NIH computing community by providing both increased capability and superior reliability to a single machine operation. Within this framework, however, the task of managing three computers rather than one presents formidable scheduling and logistical problems. Through the use of a shared input/output queue, many of the decisions currently being made manually can be automated, thus providing more immediate and accurate response to the variations in work load and resource availability.

Honors and Awards: None

Publications: None

Serial No. 2.5
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: 360 Systems Development

Previous Serial Number: None

Principal Investigator: Robert H. Brunelle

Other Investigators: Staff of Systems Team

Cooperating Units: PAL Unit

Man Years

Total:	7
Professional:	7
Others:	0

Project Description:

Objectives:

To maximize the thruput and minimize turnaround time to all users of the central facility's 360 computers. To put computing power into every researcher's lab or office thru remote terminals and software systems. In general, provide the NIH user with the best Computer Center and service he can get thru software development and hardware expansion.

Methods Employed:

Judicious selection and tailoring of software systems to the NIH environment. Acquisition and development of software and hardware to provide the researcher with the tools he needs. Constant attention to overall system software and hardware needs so that new systems and hardware are available when needed. Increasing the capacity of the system to keep ahead of the needs of the NIH users.

Significance to Program of the Division:

Constant attention to all facets of computing provides the NIH with an up-to-date computer facility catering to the needs of all NIH's researchers. Computer power in the lab brings a powerful research tool closer to the project, thus making it easier to use, and more

likely to become an integral part of, the research program. All of these efforts combine to form a forward looking Computer Center sensitive to the needs of the NIH research environment.

Honors and Awards: None

Publications: None

Serial No. 2.6
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: WYLBUR Text-editing System

Previous Serial Number: None

Principal Investigator: R. Fajman

Other Investigators: L. Barden, J. Gilman

Cooperating Units: None

Man Years

Total:	3/4
Professional:	3/4
Others:	0

Project Description:

Objectives:

1. To provide Computer Center users with a convenient and comprehensive tool to aid in the creation and development of programs.

2. To allow, secondarily, for the creation and editing of text materials such as letters, reports, etc.

3. To provide a convenient method for ascertaining the status of the computing system as a whole and locating particular jobs as they are being processed.

4. To lower the overall volume of materials which are handled by the Production Unit directly.

Methods Employed:

WYLBUR is an on-line text-editing and remote job entry system oriented to low-speed character mode terminals such as the IBM 2741, 1050, and 2260 and Models 33, 35, and 37 teletype machines. It was originally designed and implemented at the Stanford University Computation Center. WYLBUR provides the user, in his home or office, the

facility to create and edit source programs in real time, submit them for compilation and execution by the standard job stream processor, and retrieve the results of execution at his terminal. In addition, the user may work with arbitrary text material, libraries, and interrogate the system about the current status of the batch job stream.

Significance to Program of the Division:

WYLBUR represents a significant move towards lowering the amount of non-programming overhead involved in the development of programs. With the job output available at the user's terminal, effective turnaround time is substantially lowered. In addition, the added ease of using a typewriter-like terminal instead of a keypunch contributes to more rapid development and debugging of programs.

Honors and Awards: None

Publications: None

Serial No. 2.7
1. Computer Center Branch
2.
3. Betheada

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Lavender Box

Previous Serial Number: None

Principal Investigator: Robert H. Brunelle

Other Investigators: E. Alterman, J. Camp, R. Fajman, J. Gilman, W. Speary

Cooperating Units: None

Man Years

Total:	1
Professional:	1
Other:	0

Project Description:

Objectives:

NIH is acquiring a large variety of terminals that will need to communicate with the central facility. This project addresses that problem. It is an attempt to outline a plan for DCRT to establish a standard communication language that will enable any type of terminal to communicate with the central system. Further, it describes what the central facility should look like, and how and what type of services the Computer Center will be able to offer the users of remote terminals.

Methods Employed:

The plan for DCRT's implementation of such a project has been named the Lavender Box Project. The prime function of the Lavender Box Project is to provide total systems control for a large central computing complex surrounded in the periphery by multitudinous terminals. The types of terminals vary from simple typewriter terminals to small computers. The functions performed by these terminals also vary widely from conversational mode terminals, RJE, to requests for computer power or data storage.

Significance to Program of the Division:

The most important aspect of the NIH terminal system is that an NIH standard communications language be defined. Once this is done, any type of remote will be able to use the Computer Center's facilities.

Such a system will give the Computer Center, DCRT, and NIH a powerful and flexible system. The capacities of the central system will be available to remote locations, all systems will be more reliable, and the computer load will be balanced dynamically rather than manually.

Honors and Awards: None

Publications: None

Serial No. 2.8
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Data Store Program

Previous Serial Number: None

Principal Investigator: Robert H. Brunelle

Other Investigators: E. Alterman, J. Gilman, W. Speary

Cooperating Units: None

Man Years

Total:	1/3
Professional:	1/3
Others:	0

Project Description:

Objectives:

The primary purpose of the data store system is to supply the equivalent of a large on-line data and information storage capability to users of small computers outside the central computer facility. This capability will vastly increase the utility and power of the many small computers in the various laboratories and clinics.

Methods Employed:

The small remote computers will converse with the data store system via communication lines. They will request various activities to be performed with regard to a large data file. Capabilities include reading of data currently stored in the file, creating new files, duplicating files and selective processing of subsets of this data.

Significance to Program of the Division:

This is an important step in the Computer Center's plan to create an environment in which the many small specialized laboratory computers can easily converse with the powerful central facility. Effectively, it will enable the researcher to have a large computer at his finger

tips with the resultant computational and data processing power available locally to the lab, clinic or office.

Honors and Awards: None

Publications: None

Serial No. 2.9
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Pharmacology Project, AGT-30

Previous Serial Number: None

Principal Investigator: Charles R. T. Bacon

Other Investigators: Joel Kaufmann

Cooperating Units: DRFR

Man Years

Total: 2
Professional: 2
Others: 0

Project Description:

Objectives:

The system is being developed collaboratively by DRFR and DCRT as a research tool for pharmacology and toxicology. The project is a first step in DRFR's development and evaluation of an integrated set of computer techniques for handling data on chemical compounds and the functioning of living systems. DCRT is structuring its effort so that the computer systems developed can be generalized to support other areas of biomedical research and a wider community of users at NIH.

Methods Employed:

The terminal consists of a CRT display and its related imposing hardware array. It displays three-dimensional picture information with great speed and directness. In addition, it is a system for the management of pictures, and for conversational interaction with the 360 system. The terminal is a general 3-D viewing device. It displays two-dimensional projections of the three-dimensional information which is stored in its memory, and the user is free to alter viewing angles, translations, and magnifications. It will be useful to scientists with three-dimensional computation problems, e.g., crystallographers or pharmacologists can use the terminal to observe the effects of a course of calculations, and approve, modify, or cancel them using the high-speed data connection with the 360.

In the near future this system will include general purpose software in both the AGT and 360 to enable the user to manipulate his pictures and their underlying computations. Under the AGT-30 system, the picture appears as a solid object with a set of natural, easily manipulated viewing controls.

Honors and Awards: None

Publications: None

Serial No. 2.10
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Analog Data Input to the Central Computers

Previous Serial Number: None

Principal Investigator: Joseph D. Naughton

Other Investigators: David B. Gilbert, J. Rosenthal

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

The Computer Center, DCRT, is acquiring an IBM 1827 Data Control Unit which will be able to digitize up to 16 channels of analog data. This device will permit investigators with analog data to convert it to a digital form for processing by library- or user-supplied programs. The 1827 will accept analog data physically brought to the central facility on magnetic tape.

Methods Employed:

A half-inch instrumentation tape recorder will permit users with analog data on IRIG standard tapes to submit their analog data along with their jobs on a full service basis. The data tapes will be digitized and 9-track digital tapes or disk data sets will be created. These data sets will then be available for subsequent processing by the user's program.

One day turnaround on a production basis for analog data conversion is expected. The system is currently in the design phase. Testing and checkout of the equipment and programs is expected to run through the summer and analog conversion will, therefore, not be available on a service basis at that time.

The 1827 will convert analog data into a fixed point half-word format with either 8-, 11-, or 14-bit resolution. The instrumentation tape recorder will have FM electronics for all of the standard speeds between 1-7/8ips and 60ips, inclusive. In addition, direct electronics for all speeds up to 120ips will be available for one track. The first six tracks will be reserved for data, while the seventh track will contain a standard IRIG B time/event code which will be used for controlling the digitalization of the data. This unit should be in service before fall.

Significance to Program of the Division:

The ability to process analog information on the NIH central computer system will allow NIH scientists to apply the power and flexibility of the digital system to the analysis of continuous biological functions such as blood flow, blood pressure, heart beat patterns and brain waves. It will also enable them to interpret data from ultracentrifuges and radioisotope scanners without the necessity of the many intermediate and time-consuming steps presently required. This allows a great deal of clinical and research information to be analyzed which was simply ignored previously because it was impossible to process the massive amounts of information available in any reasonable amount of time.

Honors and Awards: None

Publications: None

Serial No. 2.11
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: RJE System

Previous Serial Number: None

Principal Investigator:

Other Investigators: Staff of Systems Team

Cooperating Units: None

Man Years

Total:	2
Professional:	2
Others:	0

Project Description:

Objectives:

A Remote Job Entry System (RJE) using IBM 360/20 terminal computers will allow any 360 job to be transmitted from a remote location directly to the central 360 system to be processed along with other tasks, and the results can be transmitted back to the remote location for printing. The RJE user is provided the full power of the large central 360 system at his own location including all standard programming languages, libraries, catalogued procedures, and utilities normally available in the central system when 360 jobs are submitted at the Computer Center. Use of the RJE terminals requires no programming or job control language changes. RJE terminals installed in the Institutes and Divisions will accept both express and checkout type work. There will be, however, a limit on RJE input and output volume.

Methods Employed:

The RJE terminal consists of three inter-connected units: (1) 600 card-per-minute reader; (2) a 300 line-per-minute printer with 132 print positions using a 64-character print element and a controller/transmission interface. The interface is connected to a "data set" (telephone company modulator/demodulator) which links the terminal through a standard telephone line voice grade, 2400 baud, directly to the central 360 system in Building 12.

Significance to Program of the Division:

The RJE service has had a significant impact on reducing the elapsed time between project design and actual implementation. The rapid service provided via the RJE terminal system has reduced the time required for testing and debugging new programs and algorithms from weeks or months to days or, in some cases, to mere hours. This has not merely allowed projects to be completed faster, but, because of the resultant manpower savings, has allowed many additional research programs to be conducted which would not have been possible otherwise.

Honors and Awards: None

Publications: None

Serial No. 2.12
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Reports
July 1, 1968 through June 30, 1969

Project Title: INTERFACE (technical newsletter)

Previous Serial Number: None

Principal Investigator: J. D. Naughton

Other Investigators: Staff of Systems Team and PAL Unit

Cooperating Units: None

Man Years

Total:	2
Professional:	1
Others:	1

Project Description:

Objectives:

1. The primary purpose of this effort is to give the Computer Center a regular means of conveying, (1) highly current technical information to the widely dispersed computer programmers and systems analysts at NIH, and (2) current Center accomplishments, plans, policies, course offerings and other information of interest to personnel and general managers at NIH.

2. A secondary purpose is to foster communication and collaboration among computer users and between them and the Center by providing a forum for: (1) announcing key new projects, accomplishments, personnel or organizations; and (2) airing viewpoints or suggested approaches to computing problems.

Methods Employed:

INTERFACE is published every three-to-four weeks, and distributed to all scientific and administrative personnel who have expressed a desire to be kept up to date on computing at NIH.

Significance to Program of the Division:

INTERFACE complements the Users Guide and other technical manuals by highlighting items and directing readers to the other publications for details. INTERFACE has given the users of the Computer Center a single reference point for all communications concerning the use of computers at NIH. It has done an excellent job of keeping the users informed of new services, major systems changes and all facets of computing at NIH.

Honors and Awards: None

Publications: None

Serial No. 2.13
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: IBM 2250 Graphic Program Support

Previous Serial Number: None

Principal Investigator: William Speary

Other Investigators: None

Cooperating Units: None

Man Years

Total:	1/2
Professional:	1/2
Others:	0

Project Description:

Objectives:

The objective of this project is to design, implement and maintain a PL/1 Graphic Subroutine Package to enable NIH users to communicate with the IBM 2250 Graphical Display using OS PL/1.

Methods Employed:

The PL/1 Graphic Subroutine Package is designed to enable the programm to communicate with the IBM 2250-1 using OS PL/1, version 3. The programmer using this package can have access to the 2250-1, as well as the resources of PL/1. The ability to issue calls to procedures written in assembly language enables the programmer to send data to and from the display. Interrupts are handled by user-defined ON conditions. The subroutines included in the IBM System/360 Operating System Graphic Programming Services for the IBM 2250 Display Unit, Form C27-6909, are also made accessible to the programmer.

Significance to Program of the Division:

NIH users will be able to extend the data processing power of System/360 computers: (1) to handle the graphic information

associated with medical research and analysis applications; and, (2) to provide faster and more effective retrieval and graphic expression of medical data.

Honors and Awards: None

Publications: None

Serial No. 2.14
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: CPS - Conversational Programming System

Previous Serial Number: None

Principal Investigator: John Camp

Other Investigators: Trudy Kenny

Cooperating Units: None

Man Years

Total:	1-1/2
Professional:	1-1/2
Others:	0

Project Description:

Objectives:

This is a first step in putting computing capability in the researcher's lab or office.

Methods Employed:

The CPS system allows multiple users to simultaneously write, debug and execute 360 computer programs conversationally. Programs are written in a subset of the PL/1 language from a 2741 typewriter terminal located in or near the user's office. The terminal communicates with the NIH central 360 system over telephone lines. This system is somewhat similar to the BASIC system currently available from commercial sources but has the advantage that it will be less costly to use.

Significance to Program of the Division:

CPS is the first of many steps designed to put computing capacity in the lab or office. For the first time users can communicate with the facility without leaving their natural environment. The computer is more readily available and much time is saved by not making the

trip to the central facility. Considerable computing power is placed at the user's disposal. If the CPS language is not powerful enough, the remote job entry facility places the full power of the system at his beck and call.

Honors and Awards: None

Publications: None

Serial No. 2.15

1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Biological Image Processing

Previous Serial Number: 6.2

Principal Investigators: R. J. Feldmann, S. D. Bryan

Other Investigators: A. Chauvenet, Dr. L. Lipkin

Cooperating Units: NINDS

Man Years

Total: 2.8
Professional: 2.8
Others:

Project Description:

Objectives:

1. The development of methods for man-machine interaction which permit the trained biologist to represent to the computer the relationships inherent in biological images.

2. The implementation of syntax directed picture processing techniques.

3. Research directed towards writing grammars for neuromorphology.

4. The implementation of techniques for the efficient processing of biological images.

5. The study, evaluation, and implementation of techniques for transferring biological images from one medium to another (i.e. photograph to tape, slide to tape, small computer to large computer).

Methods Employed:

A large LISP program was written to permit the on-line generation and compilation of grammars. The grammars act as the syntax for directing the processing of pictures which may be either real or abstract. Abstract pictures can be generated by drawing on the Rand Tablet of the PDP-10. Either real or abstract pictures can be presented on the

DEC 340 display. The representation of the relationships inherent in biological images is reduced to the rapid sketching of pictures which contain the relationships. The LISP program abstracts the relationships from the sketch and produces a grammar. The grammar can then be used to direct the processing of real biological images. In this manner a non-computer oriented biologist can rapidly transfer his knowledge of the relationships in biological objects to the computer.

A preliminary neuromorphological grammar has been written in conjunction with Dr. Lewis Lipkin of NINDS. The neuron grammar has been processed by the grammar compiling program, and can be used to recognize classes of abstract pictures. A simplified version of this grammar can be used for identification (not structural analysis) of neurons and associated cells in real pictures.

The picture processing language PAX developed for NIH at the University of Maryland has been converted from its IBM 360 implementation to a PDP-10 implementation. The conversion includes improvements which result in increased efficiency in computer memory utilization and running time. Algorithmns for abstracting features from real pictures have been written. Routines (in both PAX and in non-PAX like versions) are available to extract connected regions of binary pictures as well as object boundaries. Extensions to the original PAX package include spatial derivitive routines and an inverse distance transform.

Programs have been written which perform the transformation of picture data from computer to computer. Tapes which are produced by a small computer as the result of scanning photographs or slides can also be minipulated on the PDP-10 to select relevant sections.

Major Findings:

The experience gained in implementing the syntax directed picture processing system on the PDP-10 computer has indicated that increased man-machine interaction can lead to significant improvement in the ability of the computer to deal with complex problems.

Grammars can be used quite successfully to represent relationships and to direct the processing of pictures. Biological properties and relationships can be expressed in formal terms which are useful in extracting structural information from pictorial data.

The PAX package has proved to be a very useful tool in writing image processing algorithmns. Powerful programs may be developed in a very short time, an ideal combination in our experimental situation. There are inefficiencies in some PAX routines and we are rewriting the associated algorithmns in machine code.

The display and Rand Tablet facilities on the PDP-10 have proved to be extremely useful and gives us a decided advantage over many other researchers in the image processing area.

The combined use of grammars, PAX and the special I/O devices have resulted in a rapidly developing general system of considerable power.

Significance to Biomedical Research and the Program of DCRT:

The project provides the basic tools for studying the interaction of the biologist and the computer. The computer will be useful to the biologist only if it can accept and manipulate relations, concepts and objects in the same way as the biologist. The intention is to have the computer interact with the biologist in familiar terms while requiring him to formalize biological relationships.

Proposed Course of Project:

1. Gain further experience in writing grammars for a larger class of biological objects.

2. Refine the interaction between the user and the system of programs which have been implemented.

3. Increase the quality of the real-picture processing algorithmns by applying the system of programs to additional pictures.

Honors and Awards: None

Publications:

Feldmann, R. J., and Bryan, S. D.: The Interaction of Programs in a Syntax Directed Image Processing System for the PDP-10. In press.

Serial No. 2.16
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: PDP-10 Systems

Previous Serial Number: None

Principal Investigator: H. W. Vreenegoor

Other Investigators: M. C. Bruce, S. Bryan, H. Lewis, R. Feldmann

Cooperating Units: None

Man Years

Total: 3
Professional: 3
Others:

Project Description:

Objectives:

1. Development of techniques and devices to permit the use of the PDP-10 computer as a support and back-up for a network of small laboratory computers.

2. Development of software in support of a biological image processing project and heuristics efforts.

3. Establishing procedures concerning the operational use of the PDP-10 computer.

Methods Employed:

Numerous discussions were held with engineers of the Digital Equipment Corporation concerning the design and implementation of the hardware and software of the PDP-10 system. Personnel associated with the PDP-10 project received training in all aspects of the PDP-10.

Major Findings:

The PDP-10/50 computer display system was delivered and installed at DCRT in January, 1969. Acceptance testing of this system began on March 24, 1969 and was successfully completed on April 22, 1969. The

system, both hardware and software, has been functioning remarkably well with an insignificant amount of system difficulties to date. Some of the more commonly used languages such as FORTRAN, LISP, AID, and BASIC are operational in the time-shared environment.

The system has already met the needs of a variety of users with complex, atypical problems. It was possible to meet these needs rather easily since the system is designed in a modular and flexible way, permitting DCRT personnel to implement the required changes.

Significance to Biomedical Research:

As an important link in a contemplated computer network, the various NIH programs will be provided with a powerful and flexible tool as a back-up for their smaller laboratory computers. Also, since many biomedical research institutions have similar computers, we can take advantage of sophisticated software developed outside of NIH.

Additional hardware to facilitate accomplishing the stated objectives is being procured.

Detailed studies will be undertaken to learn how to best implement a data communication network.

An inexpensive, but inefficient first step in computer to computer communications is already in progress and should be completed during the coming fiscal year.

It is anticipated that a very advanced reentrant monitor will become available during the next fiscal year and will be installed by DCRT personnel.

Honors and Awards: None

Publications: None

Serial No. 2.17
1. Computer Center Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Graphics Support for PDP-10 Computer Display System

Previous Serial Number: None

Principal Investigator: Harry R. Lewis

Other Investigators: M. C. Bruce

Cooperating Units: None

Man Years:

Total: 0.5
Professional: 0.5
Others: 0.0

Project Description:

Objectives:

The goal of the PDP-10 graphics support project is to provide a convenient and efficient man-machine communications system utilizing the graphics-oriented hardware which is part of the PDP-10 computer system.

Methods Employed:

Four approaches have been used for designing and implementing a facility of the type described above:

1. Aiding in the specification and design of the special graphics hardware provided by the vendor, Digital Equipment Corporation, as part of the original computer-display system. This special hardware includes:

    a. A specially modified DEC 340 display and PDP-10-340 computer-display interface.

    b. A BBN Grafacon ("RAND tablet").

    c. A set of 16 function buttons.

    d. A high-frequency real-time clock.

2. Aiding in the specification and design of software to integrate the display hardware into the PDP-10/50 timesharing monitor.

3. Design and implementation of new software. This software is of two kinds:

   a. "Service routines," i.e., software to integrate the special hardware, other than the display, into the timesharing monitor. Only the display service routine was provided by Digital Equipment Corporation.

   b. User-oriented software, to provide a convenient facility for problem-oriented graphics work by users with minimal knowledge of computer graphics.

4. Design of new hardware to improve the operation of the existing system.

Major Findings:

1. After extensive consultation with engineers from DEC, a mutually acceptable design for the special hardware was agreed upon, one superior in several respects to the initially proposed design. After delivery of the system, several flaws in the design of the display and Grafacon and their interfaces were discovered; several weeks were devoted to diagnosing and correcting these problems before the system met the manufacturer's specifications.

2. The display service routine provided by DEC was found deficient in several respects; these problems have been corrected.

3. A large amount of new software has been produced.

   a. A service routine for the BBN Grafacon has been added to the monitor, so that this device can be used as a standard peripheral device under the timesharing monitor.

   b. Three major pieces of problem-oriented software have been written and are in general use:

      i. An assembly-level language for the DEC 340 display.

      ii. A compiler-level language for the display. This is implemented as a set of subroutines callable from a host language, which may be FORTRAN, LISP, or assembly code.

      iii. A stroke-analysis character recognizer for the input of canonical two-dimensional objects via the RAND tablet.

4. The design of a comparator has been completed. This is a device which enables the RAND tablet to simulate the action of a light pen for picking items from a display file.

Significance to Biomedical Research and the Program of DCRT:

The graphics software and hardware described above have already been used for several problems, including the display of chemical molecules and scanned images of biomedical materials. Since naive users have thus far adapted very well to the graphics system, many more projects of a graphical nature can be expected to utilize these facilities.

Proposed Course of Project:

1. Service routines for the other special hardware items are under development.

2. Construction of the comparator and programming of the necessary software to integrate it into the system will continue.

3. Further development of general-purpose graphics software will continue as the needs arise.

Honors and Awards: None

Publications: None

July 1, 1968 through June 30, 1969

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Laboratory Activities

1. DCRT-3
Serial Number

2. LABORATORY OF APPLIED STUDIES

3. Eugene K. Harris, Ph.D.
Chief

I. GENERAL SUMMARY

The primary function of this Laboratory lies in the application of mathematical statistics,mathematics and computer technology to areas of laboratory and clinical research at NIH where opportunities exist and the need is clear for productive collaboration between biomedical and mathematical or computer scientists. An important second purpose of the Laboratory is the pursuit of independent research projects, particularly in applied mathematics and computation, where these will provide valuable support to biomedical research. During the past year Laboratory staff have completed the analysis and publication, or preparation of reports of results from collaborative and independent projects in heart surgery, neural control, metabolic studies, clinical chemistry, multivariate analysis, and the mathematical theory of nucleotide sequencing. This year has also seen the initiation of new studies in such areas as the separation and analysis of multiple unit action potentials from single neuro-electrodes, the three-dimensional simulation of cardiac conduction pathways in normal and diseased states, the automation of EKG analysis, the application of mathematical methods to the determination of the primary structure of viral nucleic acids. Continuing closer collaboration with "wet" physiology laboratories in other Institutes has been achieved, permitting LAS scientists to contribute to and gain experience from on-going experimental procedures. Finally, the formal activities of this Laboratory in the education of NIH biomedical scientists to understand and use mathematical and computational methods have been considerably expanded during this year.

II. CURRENT LABORATORY PROGRAMS

A. Objectives

1. To initiate or join in sound collaborative research involving the application of mathematical statistical theory and related computer technology to problems of importance in experimental and clinical medicine.

2. To research and test new mathematical statistical methodology, mathematical models; and computational methods useful in the study of biological processes of medical significance.

3. To support these research efforts through the development of specialized computer programs to test biomathematical models and facilitate the analysis of laboratory and clinical data, often collected continuously

in analog form and requiring analog-digital conversion.

4. To support scientific studies NIH-wide through mathematical and mathematical statistical computer programming efforts.

5. To educate and train I/D scientific, technical and clinical personnel in the theory and use of mathematical and computational methods applicable to biomedical studies.

B. Summary of Current Projects

1. Collaborative Studies with NHI and NINDS.

As noted in the last Annual Report, a system of linked computer programs has been developed to analyze continuously recorded measurements of ventricular and aortic pressures, aortic flow and the corresponding EKG. Included in this programming system is an extensive array of automated graphic output. During this reporting year, a complete description and user's guide to these programs has been published and distributed as the second in a series of DCRT Technical Reports. Supported by mathematical models to describe energy losses across valvular orifices, these programs have been used in experimental studies (a) to measure operating efficiencies of ball and disc artificial valves, and (b) to evaluate the dynamic relationship between energy loss, cardiac output and estimated valve area over a single beat. These studies, initiated during the last reporting period, have now been completed (DCRT Project No. 3.5), and reports prepared for publication.

An algorithm developed within this Laboratory for detecting and diagnosing arrythmias, based on cross-correlation between EKG patterns of consecutive beats, has been implemented in real-time by the Computer Systems Laboratory, DCRT, on the hybrid computer, and is now being incorporated by CSL in equipment for use in the intensive care of cardiac patients (DCRT 4.15). Within LAS, this effort has been broadened into an extensive simulation study, in three dimensions, of the entire conduction process on a cellular level from sinoatrial node to ventricular Purkinje fiber. This will permit further understanding based on known physiological properties of the generation of normal and abnormal surface electrocardiograms (DCRT 3.6).

This year saw the beginnings of collaborative efforts with the Cardiology Branch, NHI. Of several studies currently under consideration, only one, the automated analysis of clinical electrocardiograms has progressed sufficiently to warrant reporting (DCRT 3.17). This project represents an immediate example of a more general DCRT objective, the facilitation of easy, direct access by I/D investigators to central analog-digital conversion equipment and subsequent digital computation.

Current projects in cooperation with the Surgical Neurology Branch, NINDS, are described in DCRT (3.7). A major study this year has been the examination of statistical variation and autocorrelation in neuronal discharges recorded concurrently from extracellular electrodes in the cat motor cortex (right and left) following electrical stimulation of the dentate nucleus. The major value of such studies, to date, has been the development and application of powerful computer programs for analysis of continuous and discrete time series and the separation of multiple spike trains.

Since the programming system for statistical analysis of time series has been generally available for use at NIH during the past year, this completed study (DCRT 3.9) in the last Annual Report has been omitted from this report.

New and continuing studies in collaboration with staff of the Laboratory of Neural Control and other neurophysiologists include the investigation of statistical techniques for separating multi-unit spike trains contributing to the activity recorded by a single neuroelectrode (NDS(I)-69 LNLC/OC 1687) and further study of the mathematical statistical operations of an adaptive filter for recovering variable latency evoked response from background neuroelectric activity. (DCRT 3.3)

As mentioned earlier, a particularly strong effort has been made this year to involve Laboratory scientists in ongoing experimental work of mutual interest in various I/D laboratories. One such link has been established in a study of visual processes (DCRT 3.18) in which the implications of a formal theory of visual processing developed by a LAS scientist are being explored through collaborative studies with NINDS in retinal and cortical experiments in higher and lower vertebrate forms. The work has potential for establishing a new basis for understanding the coding of visual stimuli in the retina and higher nervous structures.

2. Mathematical and Statistical Research

The application of multivariate statistical methods has always been closely allied to the use of electronic digital computing in biological work because the matrix operations required cannot normally be executed except through large computers. An important program goal of this Laboratory has been the development of expertise in this area of statistical theory to support collaborative studies in which the Laboratory is engaged. With the growth of this expertise, independently generated research projects have naturally emerged. These studies currently include the assessment of the power of various likelihood criteria to test certain assumptions underlying multivariate analysis, e.g., equality of covariance matrices (DCRT 3.11). In a related development (DCRT 3.13), a general mathematical theory defining size and shape relationships in organisms has been structured with applications, for example, to repeated measurements of children.

A study begun during the last reporting period of the application of the theory of graphs (i.e., networks) to analyse the consistency of polymer sequences following enzymatic digestion has been completed and is now in press. This project (DCRT 3.2) will be continued in combination with (DCRT 3.1) to form a collaborative study with an NIMH biochemistry laboratory in hopes of successfully applying and extending these new mathematical theories to the rigorous determination of the primary structure of nucleic acids in viruses.

A new independent study with potential applications to rapid manipulation of molecular structures on a graphics display terminal is described in (DCRT 3.8).

3. Related Studies in Application of Mathematical and Computational Methods of Biomedical Problems

Another collaboration of theory and experiment is described in (DCRT 3.20) which concerns the study and application of chemical kinetics to coupled enzyme reactions. This project is being followed in cooperation with the Clinical Pathology Department.

An independent study (DCRT 3.19) of computer-based decision strategies is being pursued to aid in the localization of neurological lesions, in the control of therapy for diabetics, and in other potential areas to assist the physician in diagnosis and therapy.

Clinical chemistry studies of normal volunteers (DCRT 3.4) have continued during this reporting year. Publication of results was delayed pending deeper analysis of accompanying serum pool data which provided estimates of analytical error. In addition, emphasis this year has been on mathematical statistical methods of separating intra-individual biological change from analytical error, in order to estimate variations in homeostatic control exerted by different individuals.

4. Mathematical and Statistical Programming Support

During this reporting period, the Laboratory assumed responsibility for the programming and analysis services provided to I/D scientists by mathematical and statistical programming staff formerly with the Computation and Data Processing Branch. The basic operating philosophy has been that of assigning personnel-many of whom have masters degrees in mathematics or statistics-to those I/D projects (including DCRT) where they have a clear opportunity to apply their academic training as well as their programming skills and to gain a coherent understanding of the biomedical concepts underlying the research project.

It is not possible in this report to give more than brief mention of some of the I/D research projects to which the programming staff have contributed. On the mathematical side, these have included

(a) the development and computer implementation of mathematical models of molecular energy functions to aid in determining structure; (b) the specification, programming and testing of an on-line graphical system for neural network modelling; (c) application and modification of a large-scale chemical equilibrium program for specifying the concentrations of blood constituents under equilibrium conditions; (d) testing of mathematical models for detecting dynamic changes in bone mineral content of living humans; (e) theoretical and numerical investigation of neuronal membrane behavior under simultaneous synaptic input and somatic action potentials; (f) preparation of a guide to the use and capabilities of the DCRT central Cal Comp plotter.

Statistical programming accomplishments during the past year have included (a) construction of information file and application of multivariate clustering techniques to analyze voluminous descriptive data on schizophrenic patients; (b) time series analysis of behavioral patterns associated with mother-child interactions; (c) determination of optimal method and time intervals for oral administration of glucose to different age groups in measuring glucose tolerance; (d) development and testing of linear programming model for application to the funding of research grant applications; (e) testing, modification, maintenance and expansion of DCRT library of mathematic and statistical computer programs.

5. Education and Training; Organizational Developments

To achieve a greater understanding by biological scientists of the applications of mathematical and computer science to biomedical research, Laboratory staff presented eight short courses on different topics, including various subjects in statistical theory; applications of graphs and networks, automotor theory and time series analysis. One staff member served on the faculty of the Foundation for Advanced Education in the Sciences, teaching a course in Elementary Probability for the Biological Sciences.

A Biomathematics and Statistics Section, headed by Dr. James E. Mosimann, has been formally established within the Laboratory.

Publications

1. Connor, R.J.: The sampling distribution of the range test for homogeneity. J. American Statistical Association. 1969 in press.

2. Connor, R.J., and Mosimann, J.E.: Concepts of independence for proportions with a generalization of the Dirichlet distribution. J. American Statistical Association. 64: 194-206, 1969.

3. Fletcher, J. and Spector, A.: A procedure for computer analysis of data from macromolecule-ligand binding studies. Computers and Biomedical Research. 2: 1968.

4. Harris, E.K. and Woody, C.D.: Use of an adaptive filter to characterize signal-noise relationships. Computers and Biomedical Research. 2: 242-273, 1969.

5. Hutchinson, G.: Evaluation of polymer sequence data using graph theory. Bull. of Math. Biophys. 1969 in press.

6. Metzger, H., Shapiro, M.B., Mosimann, J.E., and Vinton, J.E.: Assessment of compositional relatedness between proteins. Nature. 219: 1166-1168, 1968.

7. Mosimann, J.E., Jolicoeur, P.: Intervalles de confiance pour la pente de l'axe majeur d'une distribution normale bidimensionnelle. Biometrie-Praximetrie. IX: 121-140, 1968.

8. Zaveler, S.A., with Krichevsky, M.I. and Bulkeley. J.: Computer aided single or dual isotope channels ratio quench correlation in liquid scintillation counting. Analytical Biochemistry. 22: 442-464, 1968.

9. Gilbert, D.B., Digital Analysis of Hemodynamics Data, DCRT Technical Bulletin #2.

Serial No. 3.1
1. Lab. of Applied Studies
2. Biomathematics & Statistics
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Mathematical Problems of Sequence Determination of Nucleic Acids.

Previous Serial Number: Same

Principal Investigators: Carl Merril, M.D., NIMH
George Hutchinson, Ph.D., DCRT
James E. Mosimann, Ph.D., DCRT

Cooperating Units: Laboratory of General and Comparative Biochemistry, NIMH

Man Years (computed for the 12 month period):

Total: 0.5
Professional: 0.5
Other: 0.0

Project Description:

Objectives and Methods:

By working in close collaboration with Dr. C. Merril in the Laboratory of General and Comparative Biochemistry, NIMH, to: (1) Study the applicability of mathematical and Statistical methods previously developed (Mosimann et al, 1966; Merril et al, 1966; Hutchinson, in press) to laboratory data on the primary sequence of the RNA bacteriophage MS2.

(2) By learning and observing the details of laboratory methods in nucleic acid sequencing, to develop new strategies for sequence determination utilizing mathematical and statistical insights.

(3) To investigate possible application of algebraic methods in studying problems in molecular biology.

Major Findings:

Mathematical properties of sequences and sequence determination have been extensively studied. Applicability to laboratory data is being evaluated. Recently developed laboratory techniques suggesting new mathematical and computer applications are being studied.

Significance to Biomedical Research:

Determination of the primary structure of nucleic acids is of extreme importance in studies of the genetic code, as well as molecular basis of life.

References:

Hutchinson, G.: Evaluation of polymer sequence data using graph theory. Bull. of Math. Biophys., to appear, 1969.

Merril, C.R., Shapiro, M.B., Bradley, D.F., Mosimann, J.E., and Vinton, J.E.: Reconstruction of protein and nucleic acid sequences V. Computer-simulated tests of various tactics for reconstructing the sequences of transfer ribonucleic acids. Biopolymers, 4: 723-735, 1963.

Mosimann, J.E., Shapiro, M.B., Merril, C.R., Bradley, D.F. and Vinton, J.E.: Reconstruction of protein and nucleic acid sequences: IV. The algebra of free monoids and the fragmentation stratagem. Bull. of Math. Biophys. 28: 235-260, 1966.

Honors and Awards: None

Publications: As cited under previous projects.

Serial No. 3.2
1. Lab. of Applied Studies
2. Biomathematics & Statistics
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Graph Theoretic Analysis of Polymer Sequencing Data obtained by Fragmentation

Previous Serial Number: 3.2

Principal Investigator: George Hutchinson, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years (computed for the 12 month period):

Total: 0.5
Professional: 0.5
Other: 0.0

Project Description:

Objectives:

To create a mathematical model of experiments in molecular biology which determine protein and nucleic acid sequences by fragmenting the polymer through complete or partial digestion, and determining the fragment sequences.

To develop mathematical criteria by which inconsistencies in experimental data can be detected, by which a polymer sequence or sequences consistent with the data can be computed, and by which all alternative possibilities can be eliminated.

Methods Employed:

Fragment sequences from an unknown protein or nucleic acid sequence are presumed known. A mathematical model using a modification of the theory of graphs (i.e. networks) was developed. A special graph is constructed from any given data configuration of a certain class. The polymer sequences consistent with the data are in one-one correspondence with Euler circuits of the constructed graph. Therefore, the constructed graph holds the complete solution of the problem in a compact form

Major Findings:

The model has been developed for the common situation in which complete digests by pancreatic ribonuclease and T1 ribonuclease are available. In many cases, partial digest data can also be fitted into the theoretical framework. All data fitted into this framework can be completely analyzed. Criteria have been developed to determine whether the data are inconsistent, to recover the polymer sequence or sequences consistent with the data, and to eliminate all alternatives inconsistent with the data. Efficient algorithms are available for these computations.

Significance to Biomedical Research and the Program of the Institute:

The method developed can aid researchers in molecular biology to quickly and accurately evaluate data, to detect erroneous data, and to avoid redundant laboratory effort.

Proposed Course of Study:

Project was combined with project previously number 3.1

Honors and Awards: None

Publications:

Hutchinson, G., Evaluation of Polymer Sequence Data Using Graph Theory, Bull. of Math. Biophys., 1969 in press

Serial No. 3.3
1. Lab. of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Signal-Noise Relationships in Neuroelectric Data

Previous Serial Number: 3.16

Principal Investigators: Eugene K. Harris, Ph.D.
William D. Vincent

Cooperating Units: None

Man Years (computed for the 12 month period):

Total:	0.6
Professional:	0.6
Other:	0.0

Project Description:

Objectives:

To characterize band-limited noise components of neuroelectric signals and, in particular, to distinguish signal-associated from signal-unassociated components of such noise, with expectation that this discrimination will aid in understanding the statistical organization of adaptive neural systems.

Methods Employed:

Refer to 1968 Annual Report for discussion of methods and past findings. These have been described in a publication listed below (Harris and Woody). During this fiscal year, the adaptive filter technique has been extended to artificial variable latency signals, where latencies have been drawn randomly from uniform exponential, gamma and normal distributions. Such signals are being filtered, first in the absence of noise, then with added noise in increasing amounts.

Major Findings:

Initial difficulties were encountered in obtaining random variable latencies. These have only recently been overcome, but preliminary results show not only that the adaptive filter operates efficiently, but in addition appears to produce recovered signals by minimizing absolute deviations in latencies.

Significance to Biomedical Research:

These results provide simple, useful tools for describing dynamic changes in the organization of specific neural systems as monitored by extracellular deep electrodes during processes of sensory stimulation, habituation and conditioning.

Proposed Course of Project:

Computer studies of artificial variable-latency signals to determine the feasibility of estimating the distributing of latencies by application of mathematical models to output from adaptive filtering.

Honors and Awards: None

Publications:

Woody, C.D.: Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. and Biol. Eng. 5: 539-553, 1967.

Harris, E.K. and Woody, C.D: Use of an adaptive filter to characterize signal-noise relationships. Computers and Biomedical Research. 2: 242-273, 1969.

Serial No. 3.4
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Statistical Research in Clinical Pathology

Previous Serial Number: Same

Principal Investigator: Eugene K. Harris, Ph.D.

Co-Investigators: Ernest Cotlove, M.D., Clinical Pathology Dept., CC
Donald Young, M.D., Clinical Pathology Dept., CC
George Shakarji, LAS, DCRT

Man Years (computed for the 12 month period):

Total: .85
Professional: .75
Other: .10

Project Description:

Objectives:

To explore, through careful application of statistical methodology, relationships inherent in large volumes of results of chemical analyses routinely performed on serial blood samples from normal volunteers and patients at the Clinical Center.

Methods Employed:

At this time, blood chemistries from 78 normal volunteers, in ten groups of seven or eight individuals per group, are stored in easily retrievable form on digital magnetic tape. These data consist of concentrations of at least 15 elements and compounds (including some enzymes) measured in each of 10-12 weekly blood samples from each subject. The total ten-group data base covers a 2 1/2 year period of study, and includes approximately 26,000 measurements.

Similarly stored patient blood data are available from approximately 2500 Clinical Center admissions during a 9-month period in 1964-65. Demographic and diagnostic information on these patients is also available for related study. Serial blood data for each patient is less abundant than for normal volunteers and not regularly spaced, but does permit some time-wise analyses relative to effects of therapy or other changes of state.

A separate component of variance analysis has been applied by computer program to the weekly concentrations for each normal volunteer. The availability of daily measurements on pooled frozen serum allowed separate estimation of analytical error, so that weekly variation in a subject's results can be apportioned to (a) inherent physiological variability, (b) fluctuations in routine laboratory analysis of replicate samples. Analysis of individual variance components. New mathematical statistical methods have been developed to characterize the distribution of intra-individual biological variances across a population of normal individuals.

Major Findings:

New results during the past year based on more extensive analysis of serum pools, has shown the latter to provide measures of analytical error truly representative of those occurring during the analysis of patient samples. Publication of overall statistical analysis, scheduled for this year, had been delayed until this important point was fully clarified. Revised final reports are now in preparation.

Significance to Biomedical Research:

These studies will (1) add substantially to existing information on physiological variation of blood constituents within a single healthy individual over time and among normal individuals within a given age, sex, and race group, and (2) provide new knowledge on currently achievable analytical precision (e.g., in the course of multiphasic screening programs), and (3) provide an opportunity to test potentially useful mathematical models of time-dependent changes in blood constituents of patients during the course of therapy.

Proposed Course of Project:

Further analysis of data from normal subjects will study possible inter-relationships among different blood constituents. One specific aim will be the development of a function of several variables which may provide a distinctive blood "profile" of a normal individual. Later, attention will be centered on the patient file, using distributions of blood tests in normal volunteers to aid in isolating extreme ranges in each variable among patients. Correlations among extremes and relationships to diagnosis and to demographic variables will be examined, and relevant physiological hypotheses will be tested. Finally, the variation in each test over time will be studied in relation to early and final diagnoses and course of therapy. At this point, more sophisticated mathematical statistical methods, drawn from the theory of stochastic processes, will be needed to describe dynamic changes in test results during hospitalization.

Honors and Awards: None

Publications: None

Serial No. 3.5
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Heart Surgical Suite Project

Previous Serial Number: Same

Principal Investigator: David B. Gilbert, M.D.

Other Investigators: Martha R. Horton

Cooperating Units: None

Man Years (computed for the 12 month period):

Total: 1.0
Professional: 1.0
Other: 0.0

Project Description:

Objectives:

The goals of the Heart Surgical Suite Project are to provide accurate, well-organized mathematical analyses of massive quantities of physiological data obtained from humans during surgery.

Methods Employed:

Data acquired in analog form on magnetic tape during surgery are edited off-line on equipment designed by the Division. The data are then converted to digital form on the CDC 3100. The digital tape is analyzed off-line on the IBM 360 system utilizing FORTRAN and assembly language programs. The parameters presently include two pressures, the flow between these two pressures, and the electrocardiogram. Classical analyses of this data include the maxima and minima of flow and pressure, rates of flow and pressure change, stroke volume, stroke work and power. Experimental analysis includes the energy loss across a diseased valve expressed as valvular efficiency and the distribution of the fractional energy loss throughout a cycle. Reverse flow in the aorta is also calculated.

Major Findings:

The Heart Surgical Suite routines were used to analyze beat-to-beat data in the following two instances (1) Two specific types of prosthetic valves were evaluated in calves in the mitral position. Analysis of the data provided insight into the operating characteristics of each in an experimental design which utilized heart rate and stroke volumes as independent variables. The efficiency of the disc valves was heart rate dependent where as the ball valve was not. (2) The relationship between beat energy loss, cardiac output and estimated valve area was defined for two patients with left ventricular outflow obstruction. The mathematically defined relationship between these three variables emphasizes the necessity of using both valve area and cardiac output to quantitate cardiac detriment due to outflow obstruction. Evaluation of the valve area alone may lead to misinterpretation of catheterization data.

Significance to Biomedical Research:

Application of the Surgical Suite Program techniques has provided quantitation and insight into the description of valvular heart disease. The resultant mathematical models allow simplified description of the interaction of multiple variables, multiple beat analysis has allowed a definition of steady-state and beat-to-beat variation of that state.

Proposed Course of Project:

Indefinite. To be provided by subsequent investigators.

Honors and Awards: None

Publications:

Gilbert, D.B., Digital Analysis of Hemodynamics Data, DCRT Technical Bulletin #2.

Serial No. 3.6
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer Modeling of Cardiac Conduction

Previous Serial Number: Same

Principal Investigators: Martha Horton
David B. Gilbert, M.D.

Other Investigators: None

Cooperating Units: None

Man Years (computed for the 12 month period):

Total:	0.0
Professional:	0.0
Other:	0.0

Project Description:

Objectives:

To model the conduction process from sinoatrial node to ventricular Purkinje fiber utilizing known estimations of conduction velocities and action potentials of specific cell types. The resultant model attempts to predict the surface electrocardiogram given variations in the shape of cellular action potentials, disturbances in conduction or changes in the conduction velocity of specific areas.

Methods Employed:

A digital computer (IBM 360/50) is programmed such that a given number of memory locations represent, anatomically, a cluster of cardiac cells. The time of discharge of each of these cells are dependent upon the conduction velocity for that cell type. A single pass through the memory locations determines whether an individual cell is capable of firing or whether the discharging impulse has reached the cell. After firing, the cellular transmembrane action potential voltage is calculated by referring to a "template" characteristic of that cell type. That transmembrane potential is swapped out onto disk storage and is subsequently used to calculate the estimated gross vector current produced by that cell,

The resultant output then demonstrates: (1) a graphical estimation of the time of firing of a specific cardiac cell; (2) an estimation of the summed vector currents as a function of time.

Major Findings:

The estimated firing times predicted to occur on the cardiac endo and epicardium are quite close to those found in the experimental laboratory. Preliminary evidence suggests that the summed vector currents generated quite closely resemble the clinical vectorcardiogram.

Significance to Biomedical Research:

Previous attempts to characterise the "cardiac generator" have been convenient mathematical (dipole, quadrapole, ...multipole generators) approximations or easily generated engineering functions (step functions, delta functions). While the predictability of these models is reasonable, the basis for the model is not physiologic and produces little new insight into the physiological forces that govern the generation of the surface EKG or VCG. Further, none of the previous attempts have adequately modeled the T-wave in conjunction with the QRS; no attempts have been made to integrate an entire model that generates both the atrial and the ventricular portion of the surface EKG. A unified model utilizing the fundamental properties of cardiac cellular membranes (sodium and potassium currents producing action potential templates) for both the atria, a-v node, bundle, and ventricular Purkinje fibers should allow a great deal of insight into the logic that governs the generation of normal and abnormal surface electrocardiogram.

Honors and Awards: None

Publications: None

Serial No. 3.7
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Application of Advanced Mathematical Techniques in Surgical Neurology

Previous Serial Number: Same

Principal Investigator: Raymond Mejia

Co-Investigator: Maitland Baldwin, M.D.

Other Investigator: R. Minker, CCB, D. Ashbrook and M. Douglas, LAS, DCRT.

Cooperating Units: Surgical Neurology Branch, NINDS
Technical Development Section, NIMH

Man Years (computed for the 12 month period):

Total: 2.0
Professional: 1.5
Other (including computer time): 0.5

Project Description:

Objectives:

The objective is to study physiological parameters available during surgery through the application of advanced digital processing techniques. Experiments using lower primates have been conducted, as well as studies of single-cell activity in the cat.

Methods Employed:

Programs for the IBM 360 computer have been developed and applied to quantify slowly changing variables (such as temperatures), continuously changing phenomena (such as EEG) and discrete variables (pulse rate, for example).

Computer programs for the analysis of neuronal discharges have been written for the IBM 360 computer. Action potentials are treated as discrete events occuring in time, and statistics are calculated for one or more neurons.

Major Findings:

A study of electroencephalographic changes due to subconcussive and concussive blows was completed in collaboration with Drs. A. Ommaya and D. Gainsburg and reported in [NDB(1)-62 SN/0C907(c)].

A study of the statistical variation of neuronal discharges in the cat cortex due to electrical stimulation of the dentate was completed in collaboration with Drs. C.L. Li and R. Rocheson, NINDS. A rate increase due to contralateral stimulation and a predilection for certain frequencies (e.g. 20 and 60 ms intervals between discharges) were observed.

Significance to Biomedical Research and the Program of the Division:

In addition to the contribution to neurology and the art of surgery, the quantification of meaningful physiological parameters in a manner amenable to immediate as well as long term scientific study is significant. The development and application of suitable mathematical techniques of analysis is made possible in large part through the availability of adequate recording and computing equipment.

The ability to study statistically the behaviour of single cells and correlate the behaviour of a number of cells may add insight to the function of specific structures within the brain. The ability to apply large scale computational technique in the analysis neuroelectric signals will now be extended to the laboratory investigator.

Proposed Course of Project:

A study of the relationship between hemispheral and motor dominance (previously initiated but postponed due to lack of needed data) will continue. It is planned to investigate the relationship of hemispheral dominance to rate of cooling during hypothermia. The study of chimpanzees subjected to mesial cerebral incision will continue and combined with the effects of mesial spinal cord incision will also be studied.

A computer program will be implemented to separate the discharges of many neurons (by threshold detection) and record for each neuron the time between discharges. This program will be able to routinely retrieve data recorded on analog tape using an EAI 8900 hybrid computer. A generalized program written for the IBM 360 computer at DCRT will compute selected statistics for each neuron or pair of neurons. A collaborative study of evoked responses in the visual cortex of the cat will follow.

Honors and Awards: None

Publications: None

Serial No. 3.8
1. Lab. of Applied Studies
2. Biomathematics & Statistics
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Generation of Stick Figures on a Computer Graphics Display

Previous Serial Number: None

Principal Investigator: George Hutchinson, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years (computed for the 12 month period)

Total:	0.3
Professional:	0.3
Other:	0.0

Project Description:

Objectives:

To develop methods by which many stick figures can be generated at a computer graphics terminal without requiring large data inputs.

Methods Employed:

Mathematical operations were defined which produce complex stick figures from simpler ones, under specified rules. Starting with a single point, a succession of operations can quickly produce complex stick figures. It is not necessary to give point by point and line by line descriptions of the figures, since these descriptions are automatically generated by the specified sequence of operations.

Major Findings:

A computer program utilizing the IBM 2250 graphics display terminal was programmed, coded and tested. The program implemented a small number of operations for generating stick figures. It demonstrated that a large variety of stick figures could be easily generated, especially those with regular or symmetrical structures.

Significance to Biomedical Research and the Program of the Institute:

Applications involving molecular structure displays nerve networks, computer science, etc. may require generation of complex stick figures at a graphics terminal. A general purpose program for stick figure generation could relieve such users of programming effort and minimize the need to prepare input data.

Honors and Awards: None

Publications: None

Serial No. 3.11
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer Simulation Methods and Related Statistical Methodology

Previous Serial Number: Same

Principal Investigators: Richard L. Greenstreet
Robert J. Connor, Dr. Eng.

Other Investigators: David VanSant
Ray Danner

Man Years (computed for th 12 month period):

Total: .25
Professional: .25
Other:

Project Description:

Objectives:

To provide NIH with multivariate methods specifically to study the power of certain statistics used to test equality of covariance matrices from several multivariate normal populations.

Major Findings:

The statistics $-2\rho_1 \ln \lambda$, $-2\rho_2 \ln W$, -2 and $-2 \ln W$, where λ is the maximum likelihood criterion and W is Bartlett's modification, were studied. The results indicate the first two statistics achieve a significance level close to that desired, whereas the significance level for the latter two statistics is somewhat higher than expected unless the samples are large.

The power of these statistics increases as the sample size increases, or as the dimensionality of the random vector increases or as the covariance matrices are more unequal. Furthermore, power is seen to be a convex function of the number of populations.

Significance to Biomedical Research:

Multivariate techniques are used frequently in biomedical research particularly in psychometric areas. An understanding of the properties of multivariate techniques in their application.

Honors and Awards: None

Publications: None

Serial No. 3.13
1. Lab. of Applied Studies
2. Biomathematics & Statistics
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: The Statistical Analysis of Percentage or Proportional Data

Previous Serial Number: 3.13

Principal Investigator: James E. Mosimann, Ph.D.

Other Investigators: Robert J. Connor, Dr. Eng. (thru 9-68)

Cooperating Units: None

Man Years (computed for the 12 month period):

Total:	0.9
Professional:	0.7
Other:	0.2

Project Description:

Objectives:

To provide mathematical and statistical models to permit the analysis of data which are proportioned, e.g. percent serum protein constituents of blood. To provide mathematical support for biological studies of growth and differential growth. The primary mathematical methods involved are those of probability and mathematical statistics along the lines of the multivariate beta or dirichlet model (Mosimann, 1962-63).

Methods Employed: See objectives

Major Findings:

Definitions of independence for proportions have been developed, and a general class of multivariate distributions studied. Data on bone composition in rats and scute changes in turtles have been analyzed. The methods are applicable to percentage data in general. The results are in press (see publications). This phase is now terminated.

In the next phase of the study, general definitions for random size and shape variables have been developed. These provide a more general basis for study of size-correlated shape changes in organisms

than has heretofore been available. In particular, the importance of precise definitions of size is indicated. A paper by Mosimann entitled "Size Allometry: Size and Shape Variables with Characterization of the Lognormal and Generalization Gamma Distribution" has been submitted for publication.

References:

On the compound multinomial distribution, the multivariate beta distribution and correlations among proportions. Biometrika 49, No. 1-2, 65-82, 1962

On the compound negative multinomial distributior and correlations among inversely sampled pollen counts. Biometrika 47-54, 1963.

Honors and Awards: None

Publications:

Connor, R.J., Mosimann, J.E.: Concepts of independence for proportions with a generalization of the Dirichlet distribution. J. American Statistical Association. 64: 194-206, 1969.

Serial No. 3.17
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Analog to Digital Conversion on the Central Computers

Previous Serial Number: None

Principal Investigator: Robert H. Greenfield

Other Investigators: David B. Gilbert, M.D.

Cooperating Units:

Computer Center Branch, DCRT
Computer Systems Laboratory, DCRT
Cardiology Branch, NHI

Man Years:

Total:	0.9
Professional:	0.9
Other:	0.0

Project Description:

Objectives:

The goal of this project is to provide a full service analog to digital conversion facility at the Central Computer Center of DCRT for biological investigators. The immediate goal is to provide automatic computer analysis of clinical electrocardiograms on a service basis.

Methods Employed:

An IBM 1827 Data Control Unit will be used to convert the analog signals to a digital form, control their acquisition, and to present them to an IBM System 360/50 computer. An instrumentation tape with deck with a time/event code reader will be used to read analog tapes which are to be submitted for conversion to a digital form. Electrocardiographic data acquisition equipment will be used to collect and convey EKG signals to the computer for analysis. A currently available EKG analysis program has been selected to process this data.

Major Findings:

The majority of purchased and leased equipment required is either on order or has arrived. The interface is just leaving its engineering phase. The required software is currently being tailored to the NIH computer configuration.

Significance to Biomedical Research:

DCRT has recognized the need for easy entry of an analog nature into the central digital computers on a full service basis and is currently implementing a solution to this need.

Proposed Course of Project:

Construction, installation, and check out will continue.

Honors and Awards: None

Publications: None

Serial No. 3.18
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Principles of Spatial Organization in Visual Processes

Previous Serial Number: None

Principal Investigator: Harry Blum

Other Investigators: H. Wagner, M.D., NINDS
A. Norton, M.D., NINDS

Man Years:

Total: .5
Professional: .5
Other: 0

Cooperating Units: Office of Intramural Research, NINDS

Project Description:

Objectives:

To understand how the spatial distribution of light (shape) is coded and processed in the nervous system and how these distributions may be formalized for biological use as in morphology and taxonomy.

Methods Employed:

Formalizing, developing, exploring and simulating models of shape processing in the retina and cortex of vertibrates. Design of attack on the problem via experiments.

Major Findings:

Previous formal studies have been extended and generalized to the 'Generalized Medical Axis' which suggests a number of primitive differences between lower and higher vertibrate vision. This generalized formalism also provides a capability for dealing with texturally rather than edge or line defined forms. Further this extended transform may provide new techniques for data segmentation and clustering, required in statistical and taxanomic analysis.

Significance to Biomedical Research:

The work has broad potential for understanding the visual process - both retinal and cortecal, as well as providing new clues for the coding of stimuli in the central nervous system. The work also provides new ways to quantify morphous shapes useful for computer analysis and for morphology.

Proposed Course of Project:

1) To develop formal methods for describing shape.

2) To explore experimentally (physiologically and psychologically) the range of applicability of these methods to understanding vision in animals.

3) To explore applications of these method for computer processing statistical description and morphology.

4) To study new devices suggested by the aforementioned work.

Honors and Awards: None

Publications: None

Serial No. 3.19
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer-aided Diagnosis and Therapy

Previous Serial Number: None

Principal Investigator: Mordecai Koenigsberg, M.D.

Other Investigator: Lewis B. Sheiner, M.D.

Cooperating Units: Computer Systems Laboratory, DCRT

Man Years (computed for the 12 month period):

Total:	1.0
Professional:	1.0
Other:	0.0

Project Description:

Objectives:

The goals of this project are to design programs to aid physicians in clinical medical application.

Methods Employed:

Areas of clinical medicine, involving both diagnoses and therapy, that would lend themselves to computer implementation were chosen from the medical literature. The physician's strategies in dianosing and treating were then reduced logically and numerically into algorithms for programming on the computer. Where necessary appropriate mathematical models of the systems being studied were devised to enable a prediction as to the results of therapy, so that appropriate theraputic decisions could be made.

Major Findings:

A mathematical model of the blood sugar level control system was constructed to aid in studying the system dynamics of diabetic pathology. Programs were written in PL/1 to aid in determining optional insulin and diet therapy for diabetics. The model has worked for simple situations in predicting insulin dosage, but the time for numerical integration must be shortened for practicability.

In the diagnostic area, work has progressed on the formulation of strategies to aid the physician in localizing neurological lesions, both in the peripheral and central nervous systems. Various forms of data representation and different algorithms were tested. Currently, a program has been written to localize mass lesions in the peripheral nervous system and a similar program to localize both mass and vascular lesions in the central nervous system between the spinal cord and the thalamus is being completed.

Significance to Biomedical Research and the Program of the Division:

These studies will add to a growing list of computer programs for diagnoses and therapy that can be used as a practical aid to the physician. They will demonstrate the exploitation of the processing and storage capabilities of the computer to evaluate specific data to assist the physician in his task of making accurate medical diagnoses and improving therapeutic control.

Proposed Course of Project:

The programs for localizing neurological lesions will be tested with actual patient data. Eventually they will be linked together and expanded to include higher cortical areas. Physicians will be encouraged to use the programs. Further attempts will be made to simplify the mathematical model of the blood sugar systems and expand its capacity to opteurize insulin dosage and diet therapy in typical clinical situations. The program will then be tested with actual patient data. Other areas of medical application amenable to such treatment will be studied.

Honors and Awards: None

Publications: None

Serial No. 3.20
1. Lab. of Applied Studies
2. Biomedical Studies Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Enzymatic Analysis

Previous Serial Number: None

Principal Investigator: George J. Atta

Co-Investigators: Ernest Cotlove, M.D., Clinical Pathology Dept., CC
Clement J. McDonald, M.D., Clinical Pathology Dept., CC

Man Years (computed for the 12 month period):

Total: 0.5
Professional: 0.5
Other: 0.0

Project Description:

Objectives:

To assay the enzyme activity in biological fluids of patients at the Clinical Center.

Methods Employed:

The primary methods involved are those of chemical kinetics, differential equations, and non-linear curve fitting.

Major Findings:

The activity of an enzyme is defined as the maximum velocity of the chemical reaction catalyzed by the enzyme.

It is assumed that a substance A is converted to a product B with an "auxiliary enzyme. "The product B is in turn converted to a second product C with an "indicator eyzyme." The concentration of the starting material A as a function of time can be represented as a linear combination of two exponential terms. This reaction curve shows the typical S-shape characteristic of all consecutive reactions.

Preliminary analysis of patient indicates that the exponential model fits the data adequately. However, in patients under routine

assay conditions, the enzymatic reaction can only be observed over a limited time interval. Thus in some cases the complete S-shape curve and hence the inflection point cannot be obtained.

Significance to Biomedical Research:

Only rudimentary knowledge of the principles underlying enzyme activity now exists. Consecutive assays may provide a powerful technique in determinations of enzyme activities. Adequate instrumentation is now available to provide the necessary data, precision, and analysis to investigate and assess the technique. If the method is applicable, it will be valuable as a basis for standardization of enzyme activity for clinical research. Furthermore, accurate determination of enzyme activity in patients as well as normal volunteers in the Clinical Center is of considerable importance to the attending physicians.

Proposed Course of Study:

Mathematical algorithms will be developed to utilize enzyme data and to investigate the consecutive assay technique for the determination of enzyme activity in patients.

An attempt will be made to derive a model that will enable one to measure the maximum velocity (enzyme activity) of an enzamator reaction over a very limited time interval.

Honors and Awards: None

Publications: None

July 1, 1968 through June 30, 1969

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT - 4
Serial Number

2. COMPUTER SYSTEMS LABORATORY

3. Alan Demmerle
Chief

I. SUMMARY

This Laboratory, now in its third year of operation, is currently pursuing three broad objectives in support of the conviction that computer technology can make substantial contributions to biomedical research in ways other than conventional data processing. These objectives are: the development of processing systems for laboratory and clinical data, the development of new data processing techniques, and the development of new applications for computers. The projects through which this Laboratory is pursuing these objectives include: the design and development of computer systems to collect and manipulate laboratory data and control experiments in real-time, systems for off-line data collection, a hybrid (analog/digital) computing facility, components of medical information retrieval systems, and other generally new applications of computers. Most of these projects rely more on teamwork over extended periods of time than on individual effort. The design and development of computer systems, which represent a major part of our work, require from two to three years from conception to useful productivity, as well as the efforts of teams of four to six men with diverse backgrounds. Several of these long term system development efforts, started one or two years ago, are in the middle to latter half of their development period. All of these projects show great promise.

II. CURRENT LABORATORY PROGRAMS

The Computer Systems Laboratory has continued toward its previously established, system-oriented goal. It has grown slightly in size during the last year, and is now composed of 38 individuals, slightly more than one third of whom represent the discipline of engineering, and about two thirds representing the disciplines of mathematics, linguistics, and medicine. These individuals could all be more properly called "Information System Specialists." An essential ingredient to the success of many of our projects is the cooperative efforts of individuals with varied backgrounds.

A. Objectives

The Laboratory is dedicated to three broad objectives:

1. The development of processing systems for laboratory and clinical data

a. Through the use of separate, dedicated, general-purpose computer systems

b. Through data collection systems for use in conjunction with a central computing facility

c. Through the use of special-purpose, hard-wired data processors

2. Research, development, and implementation of new data processing techniques emphasizing the aspects of methodology, systems, hardware, and software.

3. Development and implementation of new applications for computers.

B. Progress

The objectives previously set forth represent the current goals of this Laboratory. Projects have been undertaken in pursuit of each of these objectives and are briefly reviewed below. The projects are discussed in greater detail in the Individual Project Reports.

Progress Toward Objective 1 (Development of processing systems for laboratory and clinical data)

There remain many project areas at NIH in which the use of a computer as a tool has not yet been explored or exploited. The laboratory and the clinical environment are particularly fertile areas in which computer technology can be applied. To date, however, our concentration has been primarily in the laboratory environment. It is our intention to help uncover project areas in both environments, and, by working in cooperation with the scientists involved, design systems to meet the data processing requirements of these project areas. The term "data processing" is to be considered in its broadest sense, and includes the collection, manipulation, computation, and display of data. These data processing requirements tend to fall into three categories:

a. Those requiring the full support of a general-purpose computer system at the site because of the size and complexity of the requirement or the number of requirements in a limited vicinity, or the requirement for "real-time" processing of data and process control.

b. Those requiring data collection (and possibly display and transmission for analysis, possibly at some later time, at some computing facility remote from the source of the data.

c. Those requiring specific processing that can be most economically implemented by hard-wired systems.

We have examined several project areas at NIH, and designed appropriate systems to meet their requirements.

1. a. We are currently involved in various stages of development of five complete computer systems to serve the laboratory environment. We have

designed and already purchased the principal components, and are currently developing three of these systems, one to serve the laboratories of NIDR, one to serve the biochemical laboratories of NIAMD in Building 2, and one to serve the mass spectrometers of the Laboratory of Metabolism of NHI (respectively described in detail in Individual Project Reports DCRT 4.1, 4.10, and 4.11). We have also designed, and are currently involved with the procurement of, two additional computer systems, one to meet the needs of the NIMH scientists in Building 10, and one for the gerontology researchers of NICHHD in Baltimore (Individual Project Reports DCRT 4.23 and 4.24). All five systems are composed, in part, of commercially available equipment and software which are purchased, and, in part, by non-available equipment and software which are being designed and implemented in this Laboratory. Our experience has shown that about one third of the cost of the system is in the purchase price of the commercially available computer, about one third in the hardware for the interconnection and communication between the computer and the laboratory instruments, including control consoles located in the laboratories and computer site preparation, and about one third for the salaries of the government information scientists who are developing the system. When implemented, these systems will represent a significant advance in the application of computation to biology and medicine. The systems will, in general, be used for collecting and analyzing data, in real-time, from instruments directly connected to the computer; controlling experiments while they are being conducted; and developing new ways to do experiments which become possible only by virtue of the computational power of the computer. In addition, of course, these systems will serve the traditional function of analyzing experimental data that has been collected.

This Laboratory has assumed primary responsibility for achieving successful system operation. This involves, in addition to system design, the development of the interfaces between the instruments and the computer, implementing remote communication links, and developing the applications and system controlling software. The systems will then, for the most part, be turned over to the scientists who expect to use them. In the final analysis, the value of these systems depends upon the laboratory scientists learning to use the computer as a tool, since a fundamental appreciation of its capability is necessary if it is to continue to be used in an innovative way.

In addition to the actual design of computer systems for use at NIH, we have provided consultation on the design and use of computer systems for extramural programs. The largest user of our expertise in this area was the NHI, whom we have helped on a continuing basis, particularly with the data management portion of the Myocardial Infarction Research Unit (MIRU) program. During the past year, we have designed a computer system suitable for use by any of the MIRU units. If several of the MIRU units would use this system, the Government would benefit by the economies of quantity purchase, and, most important, all concerned would benefit by a one-time development of software and hardware systems. This is described in greater detail in Individual Project Report DCRT 4.12.

1. b. Complete computer systems are not always appropriate to meet the data processing requirements of laboratories. Often, collection of data on digital magnetic tape, analog magnetic tape, a punched paper tape, or direct transmission of the data for analysis at a central computer facility is the best solution. Work in this area is reported in Individual Project Reports DCRT 4.13 and 4.6.

For general use with such off-line data recording systems, we have developed time and event code generators and decoders. These allow the user of an automated data collection system to keep better track of the relationship between his experiment and his data. The generator is used at the source of the data recording, and the decoder is used at the site of the data processing. These units are described in greater detail in Individual Project Report DCRT 4.14.

1. c. Data processing requirements can sometimes be best met by special purpose computing systems, as opposed to the general purpose computers that we think of when we hear the word "computer." The special purpose systems, as the name implies, are designed for a specific job, and the logic is generally wired, although sometimes through a patchboard, so that they are not as easily adapted to other jobs as a stored program general purpose computer. They have a definite economic advantage when a single well-defined function is to be done repetitively. There are many uses for such machines at NIH. The monitoring of critically ill patients is one such area where these machines will be used in conjunction with general purpose machines. This Laboratory is currently developing such a special purpose computer for ECG analysis. The algorithms for the analysis were developed on the Laboratory's Hybrid Computer System. Special purpose hardware, using these software algorithms, was then designed, and it provides for a faster and more economical beat-to-beat ECG analysis than would be possible with a general purpose computer system. The end result is small, portable equipment for continuous real-time analysis of ECG data at the bedside. This is described in greater detail in the Individual Project Report DCRT 4.15.

<u>Progress Toward Objective 2</u> (Research, development, and implementation of new data <u>processing techniques</u>)

During this reporting period, work directed toward this objective has been oriented more toward specific requirements than in the previous year. Work toward developing new methodologies and hardware and software techniques was done, but in conjunction with, and directed toward, the data processing systems reported under our Objective 1. For example, the NIDR and NIAMD computer systems, we believe, will be especially useful if we can develop a methodology whereby the system is, in effect, an extension of the experimenter. Close and versatile two-way communication between the system and the user is the essence of this concept. We have just begun work in this area, and, if successful, it will represent a substantial contribution to the computer science. These systems also require complex system software (or monitors) to control the computer system itself in the acquisition, manipulation, and display of experimental data While some software of this type is available commercially for these particular computers, that which is available is not totally suited to our needs. It

requires too much computer core memory and too much computer time to execute, primarily because it was designed to satisfy the general user with simple applications rather than optimally meet complex, stringent requirements, such as those of the NIDR and NIAMD applications. Hence, we have done considerable work in the development of real-time, time-shared monitors which are necessary for these systems. In addition, control consoles, interfacing equipment and data display hardware have been developed in this Laboratory for these specific computer systems.

The Hybrid Computer System (located in Building 10) has provided good service to a large number of users during the past year. It is used for hybrid computation, analog computation, digital computation, and analog to digital conversion. It is used mostly for standard digital computation, and not very much in the hybrid mode where it can provide a powerful tool for the solution of biomedical problems which involve biological systems modeling and simulation. Perhaps its most useful feature is that it allows for "real-time" experiments which require a computer in conjunction with signal preprocessing equipment. The elements of the analog computer can be used for various kinds of signal processing. It has gotten heavy use in this kind of work, not only by this Laboratory, but by others in DCRT and NIH. The facility has been improved during the year by the addition of disks and a higher speed printer. This project is more fully discussed in Individual Project Report DCRT 4.9.

<u>Progress Toward Objective 3</u> (Development and implementation of new applications for computers)

This objective, which overlaps Objective 1 because of its broadness, is an extremely important part of the mission of this Laboratory and Division. We have pursued this goal through several projects. We are continuing to work on a project which will lead to both the automatic abstracting of textual data and to better information retrieval systems. Key information, abstracted with the use of a computer doing linguistic analysis of autopsy reports, currently provides an input data source for a pathology data retrieval system. It will be extended for use in information retrieval systems using other types of medical textual data. The expectations and methods employed in this project are more thoroughly described in Individual Project Report DCRT 4.19. A related project in Computer Indexing of Medical Abstracts is described in Individual Project Report DCRT 4.7.

Another project which holds great promise is one for the purpose of developing the methodology and technology whereby the potential of the computer is made available in the physician's office at a reasonable cost. Such a system would permit access to the memory and computational abilities of a computer for use in a number of applications, such as to aid in diagnosis or to help compute treatment protocols. The essence of this project is to provide computer service that is easy to use, highly reliable, and reasonably inexpensive. We have concentrated on perfecting the touch-tone telephone as an input device, and a computer driven voice response as an output, as well as the development of useful medical application programs for use with such a service. We are currently expecting to procure a computer system to serve as a communication controller

in a network which will include the NIH central facility and many of the commercially available time-shared computer services, and to serve as the base of the voice response for this system. This fascinating application is discussed more fully in Individual Project Reports DCRT 4.20 and 4.3.

Other applications of computation which have been pursued primarily through collaboration with other Institutes include the Analysis of Laboratory Spectral Data; the Sequence Determination of Proteins and Nucleic Acids; the Analysis of Wave Propagation in Arteries; Psycholinguistic Analysis; and some work involving the small computer, called the Programmed Console, used in Radiation Therapy in the Clinical Center. Work was also done in the development of a file structure for medical record information. Detailed information on these projects is to be found in Individual Project Reports DCRT 4.21, 4.5, 4.26, 4.2, 4.8, and 4.18, respectively.

General Remarks

There are a number of other projects, particularly in the clinical area, which could, and should, be undertaken, and which have the moral support of the management involved. They will probably have to wait for the completion of our current projects because the availability of additional personnel resources seems unlikely. Many of these projects cannot be undertaken half-heartedly because they require a substantial effort over an extended period of time to assure their success.

This Laboratory is appreciative of the support which has been made available to us for the execution of the projects reported. We have had adequate financial support and adequate personnel resources, and, most important, the enthusiastic support from the management involved.

Serial No. DCRT - 4.1
1. Computer Systems Laboratory
2. Systems Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Dental Institute Computing System

Previous Serial Number: 4.1

Principal Investigator: Daniel Syed

Other Investigators: Leroy Brown, William Holsinger, John Knight, William Risso

Cooperating Units: National Institute of Dental Research

Man Years:

Total: 4
Professional: 4
Other: 0

Project Description:

Objectives:

This project is intended to provide on-line real-time computational support for the biochemistry and neurophysiology research programs of the National Institute of Dental Research. The computer system is designed to automate data acquisition from instruments in the biochemistry laboratories, to perform limited computations on captured data, and, eventually, to control experiments. During neurophysiological research procedures, the system will provide information relative to the proper placement of electrodes, pertinent waveforms enhanced by averaging techniques to improve signal-to-noise ratios, and correlative relationships. The system will provide a test bed for the development of hardware and software techniques for automating many instruments that have not been automated on even an off-line basis.

Methods Employed:

A Honeywell DDP-516 computer has been procured, and is expected to be installed in the Dental Institute during the month of June, 1969. Hardware interfaces for the amino acid analyzers, scintillation counters, and neurophysiology experiments have been designed and developed. Substantial additions and modifications to the real-time

monitor of the DDP-516 have been completed. The generation of programs to service the applications cited above is currently in progress. It has been decided that initial linkage with the central computer facility will be effected via emulation, in the DDP-516, of a remote job entry IBM 360/20, and the mechanics of the implementation are now under study.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

The proposed system will permit this Division to develop software and hardware for use in similar applications around the campus. Such a system will be used as a research tool in the design of an optimal system for biomedical applications stressing data acquisition. Consistent with the development of time-sharing software on the Central Computer Facility IBM 360/50, the system will provide a basis for analysis and development of satellite computing techniques in a biomedical environment.

Proposed Course of Project:

Programs and interfaces for additional instruments and increasingly complex experiments will be developed. Specifically, additional instruments scheduled for automation include gas chromatographs, an X-ray diffractometer, a machinist's comparator, a spectrophotometer and a pH monitor. The real-time requirements associated with neurophysiological experiments will be greatly expanded.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.2
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Psycholinguistic and Adaptive Patterns in Relation to Psychophysiological Response Dispositions

Previous Serial Number: None

Principal Investigator: Hanna K. Ulatowska

Other Investigators: Lyman Wynne, Winfield H. Scott

Cooperating Units: Adult Psychiatry Branch, NIMH

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

1. Study of response dispositions to the visual and speech stimuli which are discernible in the linguistic and transactional styles of a variety of subjects, both psychiatric and non-psychiatric.

2. Developing a manual of scorable features for perceptual and cognitive styles of the two groups of subjects classified as augmentors and reducers on the basis of psychophysiological tests.

3. Quantification of some of the scorable features in terms of their linguistic exponents in order to organize and compact the data and make it operational for computer analysis and storage.

Methods Employed:

Earlier work on the project, with a small group of psychiatric patients who demonstrated augmenting and reducing patterns, suggested differences between them in psycholinguistic and adaptive patterns, as revealed in Rorschach performances. The differences were discernible regardless of degree of psychopathology and personality disorganization in the subjects. The Rorschach response styles have been identified as "interpretive" and "recognitory" styles. During

the past year, scorable features of the Interpretive and Recognitory styles have been developed into scales which can be applied to Rorschach protocols. Twenty non-psychiatric volunteers from the Montgomery Junior College, who demonstrate clear augmenting and reducing patterns on average evoked responses to visual stimuli, have been psychologically tested. The scales of Interpretive and Recognitory features are being applied blindly to their Rorschach protocols to test the strength of the association between interpretive and recognitory styles on one hand and augmentation and reduction on the other. The subjects are also being administered speech perception tests in order to investigate the relationship between visual and speech perception, and the differences between the augmentors and reducers along that particular dimension.

Major Findings:

A blind evaluation of Rorschach protocols and other test material is in progress. While the scores on Interpretive and Recognitory features permit differentiation of subjects into two groups, it is not yet known what the degree of association is between interpretive and recognitory tendencies on the Rorschach and augmentation-reduction.

Significance to Mental Health Research and the Program of DCRT:

The research may ultimately allow development of new systems of classifying patients in terms of relevant diagnostic, therapeutic and prognostic variables which are associated with fundamental aspects of psychophysiological functioning. Moreover, it is hoped that the research will provide a tool for a much faster and more objective scoring of the Rorschach tests in terms of a computer scoring system for Rorschach responses.

Proposed Course of Project:

It is proposed to refine the lists of Interpretive and Recognitory features in light of the findings in the current research with normal augmentors and reducers. The scales may then be applied to a range of normal and psychiatric subjects to further study the relationship between Interpretive-Recognitory and psycholinguistic styles. Various techniques of computerizing the psychiatric data will be explored.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.3
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Medical Applications Programs

Previous Serial Number: None

Principal Investigators: Lewis Sheiner, Paul Moyer

Other Investigators: Michael Otten, Richard Blide, Martin Koenigsberg

Cooperating Units: Pulmonary Metabolism Laboratory, University of Maryland

Man Years:

Total: 1.5
Professional: 1.5
Other: 0

Project Description:

Objectives:

To provide algorithms and checked-out programs to be used by medical professionals for assistance in patient care and clinical problems. A secondary objective is the investigation of time-sharing computer facilities and voice answer-back techniques as a means of supplying computational assistance to the practicing physician.

Methods Employed:

Various algorithms were conceived by examining the medical literature for medical problems suitable for computation and logical decision processes, and by observing and discussing with physicians the aspects of patient care amenable to computer assistance. Computer programs were written and implemented on the computer. Finally, the programs were checked with patient data for logical and mathematical accuracy.

Major Findings:

Programs were written and checked for the following medical applications: intravenous drug therapy compatibility mixtures; anti-coagulant therapy dosages; antibiotic drug dosage calculations; "most likely" neurological lesion in the peripheral nervous system, given a set of

input deficits; a symptom-diagnosis program to suggest possible diagnoses for combinations of laboratory findings for certain metabolic diseases; and a program to assist in the computational and clerical effort involved in pulmonary function test analysis. Although all of these programs have been checked out for accuracy and content, their utility in medical information telecommunications will have to be tested in a real-time medical environment.

Significance to Biomedical Research and the Program of DCRT:

This project is aimed at providing an extensive library of medical applications programs to be available for selection and trial by the clinical and research physicians. It is expected that if a sufficient variety of useful programs are available, then the utility and feasibility of medical telecommunications can be evaluated.

Proposed Course of Project:

Many applications of the computer to medical problems remain to be programmed. We expect to continue efforts toward this end.

Honors and Awards: None

Publications:

Allen, S. I. and Otten, M., "The Telephone as a Computer Input/Output Terminal for Medical Information," Journal of the American Medical Association, 208: 673-679, April 1969.

Serial No. DCRT - 4.4
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Hospital Patient Record and Information Retrieval System

Previous Serial Number: 4.4

Principal Investigator: Victor Colburn

Other Investigators: William White, Michael Otten

Cooperating Units: None

Man Years:

Total: 0
Professional: 0
Other: 0

Project Description:

Objectives:

Early in fiscal 1968, the previously reported objectives of this project were shifted to include procurement and implementation of a general-purpose computer display facility. The previous objectives involved a planned approach to application of computer system technology within the NIH Clinical Center. These earlier objectives remain intact, with the intention that they will constitute a primary use of the proposed new Computer Display Facility.

Methods Employed:

This project was reported as being in a dormant state at the end of fiscal 1968, and there has been no further activity during fiscal 1969.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

The procurement by DCRT of the proposed general purpose real-time computer display facility still appears to be a worthy objective.

We are convinced that when such a facility is available, it will find much application. Many computer applications now implemented on teletype remote service will find new life in the much greater speed and flexibility of the proposed facility.

Proposed Course of Project:

This project remains in a dormant state due to lack of funds and manpower. If funds and manpower can be allocated during fiscal 1970, the project may be reactivated along previously indicated lines.

Honors and Awards: None

Publications: None

Serial No. DCRT 4.5
1. Computer Systems Laboratory
2. Laboratory Applications
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Sequence Determination of Proteins and Nucleic Acids

Previous Serial Number: 4.5

Principal Investigator: Marvin Shapiro

Other Investigators: Jay Vinton, Lewis Sheiner, Carl Merril

Cooperating Units: Laboratory of Neurochemistry, NIMH; Laboratory of General and Comparative Biochemistry, NIMH

Man Years:

Total: 0
Professional: 0
Other: 0

Project Description:

Objectives:

The mathematical reconstruction of protein and RNA sequences from laboratory data, and the computer simulation of the laboratory process of sequencing a biopolymer.

Methods Employed:

The development of an algorithm for reconstructing sequences, and the use of this algorithm in a program which simulated the steps followed in the laboratory sequencing of an RNA molecule.

Major Findings: None

(The project was inactive during the past year.)

Significance to Biomedical Research and the Program of DCRT:

Sequencing large molecules is vital to an understanding of their structure and functions. Automated methods involving computers should be of great help in speeding up the enormous amount of work involved in the problem of sequencing.

Proposed Course of Project:

Further work will be done in modifying the algorithm and the computing methods used, so that very long (> 3000 in length) sequences can be automatically reconstructed.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.6
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Analog Transmission System for Low Frequency Biomedical Signals

Previous Serial Number: None

Principal Investigator: David Songco

Other Investigators: None

Cooperating Units: None

Man Years:

Total: .3
Professional: .2
Other: .1

Project Description:

Objectives:

To develop an analog transmission system with minimum distortion for low frequency biomedical signals

Methods Employed:

The transmission system was designed to accommodate analog signals in the frequency range of 0 to 2 K Hz. with a maximum input amplitude of 20 volts peak. It consists of a filtered input driver, twisted pair shielded transmission line, and a filtered differential line receiver.

Major Findings:

The system, as installed in the Clinical Center, using approximately 150 feet of cable, is now in use for transmission of mass spectrometer data to a computer, from a spectrometer at a remote location (as described in DCRT 4.11). The transmission characteristics of the system preserve signal amplitude and phase to within 0.5%. This fidélity exceeds source instrument accuracy.

Significance to Biomedical Research and the Program of DCRT:

When several devices share a central processor, it is often not feasible to locate them in close proximity. Analog signal transmission is one solution to this problem and requires only minimal hardware at each device.

Proposed Course of Project:

The techniques developed for this system have provided the basis for similar analog transmission configurations which will be used in conjunction with the computer system soon to be installed for NIAMD.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.7
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer Indexing of Medical Abstracts

Previous Serial Number: None

Principal Investigator: Scott Allen

Other Investigators: Milos Pacak, George Dunham

Cooperating Units: None

Man Years:

Total: .5
Professional: .5
Other: 0

Project Description:

Objectives:

To develop techniques and computer programs for the analysis or retrieval of disease description abstracts and therapeutic recommendations which are contained in current medical literature.

Methods Employed:

Key-word-in-context indexing programs were developed as an initial tool to process the 3,000 current medical terminology (CMT) abstracts proposed by the American Medical Association. Batch and on-line programs are currently operational for the extraction of subsets from the CMT file. This literature source was also tested with the Systematized Nomenclature of Pathology (SNOP) encoder programs developed in this Laboratory.

Major Findings:

Small subsets of the CMT have been encoded in a uniform structure for computer query, using pushbutton telephones and audio-response output. Additional clinical terminology has been identified for possible inclusion in the SNOP dictionary.

Significance to Biomedical Research and the Program of DCRT:

Basic concepts, thesauri, and programs developed will be useful in literature indexing and patient record processing by machine.

Proposed Course of Project:

The development of a computer-based thesauris of clinical signs, symptoms, and laboratory tests is underway. Extensive syntactic and semantic analysis is required to handle the complex array of medical terms, including prefixes, suffixes, abbreviations, idioms, and adjectival qualifiers.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.8
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Support of Programmed Console

Previous Serial Number: 4.8

Principal Investigator: Victor Colburn

Other Investigators: William White, Robert Brunelle

Cooperating Units: Radiation Branch, NCI

Man Years:

Total: .2
Professional: .2
Other: 0

Project Description:

Objectives:

To provide support for the Spear Programmed Console in the Radiation Branch of the NCI by expanding and upgrading the hardware/software system and operating techniques so as to increase its overall effectiveness.

Methods Employed:

Plans have been completed for connecting the Programmed Console, (PC) to the IBM 360/50 Computer System in the NIH Computer Center. This connection will provide much-needed additional file storage and computational assistance for the PC. A similar connection has been implemented at Washington University School of Medicine, St. Louis, Missouri, and we have drawn from their experience.

Major Findings:

The computer-to-computer connection will provide the PC with IBM 360 disc access, but will not initially provide direct entry into the 360/50 Computer System job stream. Data stored on the 360/50 System disc will be accessed and manipulated as required from a remote job entry terminal. The new PC connection will provide full and effective use of the PC/360 assembler program previously written by this Laboratory.

Significance to Biomedical Research and the Program of DCRT:

The Spear Programmed Console, as an independent machine, has been established as an essential tool for the derivation of radiation dose plans for therapeutic applications in the NCI. Connection of the PC to the 360/50 computer system will multiply its effectiveness several fold by providing:

1. Greater throughput by means of simplified operating procedures.

2. Wider application and sophistication of technique through more extensive computational capability.

This upgrading of PC system capabilities will permit vital medical research on radiation therapy to proceed at a significantly greater pace, and, at the same time, will further the interests of DCRT in extending the powerful computational services of the central NIH Computer Center to other parts of the Campus. The Spear Programmed Console will be the first small computer of its class to be so connected with the central NIH facility.

Proposed Course of Project:

As the link from the PC to the 360/50 computer system becomes operational, it is proposed that a 2741 Communications Terminal be installed near the Programmed Console to provide ready access to the 360/50 system for manipulation and control of PC stored data. Eventually, the 360/50 computer system should be further extended to permit direct access to the job stream from the Programmed Console.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.9
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Hybrid Computer Project

Previous Serial Number: 4.9

Principal Investigator: Perry S. Plexico

Other Investigators: Leroy Brown, James Del Priore, Robert Romanoff, Philip Turner

Cooperating Units: None

Man Years:

Total: 3.6
Professional: 2.9
Other: .7

Project Description:

Objectives:

1. Continuing development of methods of applying the hybrid computer to the simulation and modeling of biological systems.

2. Improvement of existing methods for data reduction, editing, and conversion.

3. Collaboration with, and support of, investigators using the computer for a variety of real-time computing and data processing tasks.

Methods Employed:

Expansion of the system to include mass storage capability was accomplished to improve the efficiency of the system in performing both real-time and batch processing tasks. Substantial additional software has been developed to enable the use of FORTRAN for real-time programming and to permit priority interrupt handling by FORTRAN programs.

Major Findings:

As expected, addition of mass storage and its accompanying operating system and other real-time software has resulted in a substantial increase in the number and variety of projects being supported by the hybrid computer.

Existing analog tape editing hardware and procedures were shown to be inadequate to meet increasing data reduction and conversion requirements. A new time and event coding and reading facility has been designed (described in detail in Individual Project Report DCRT 4.14) and will be interfaced to the hybrid computer to alleviate this situation.

Significance to Biomedical Research and the Program of DCRT:

The hybrid computer offers the only data conversion and real-time data analysis facility available on an NIH-wide basis. Continuing software and hardware development has substantially improved system utility in these areas.

Investigators from several Institutes and Divisions have been informally trained in the use of the Hybrid System, and have applied it to a variety of research and scientific computation tasks (see, for example, Individual Project Reports DCRT 4.26 and DCRT 4.20).

Utilization of the system by these individuals, representing seventeen Laboratories and Branches, has been such that the digital computer is now consistently used in excess of 90% of a single shift The analog computer, although used more sporadically, is operated, on the average, 40% of a single shift. In addition, the system is frequently used by DCRT staff members for weekend and evening work.

Proposed Course of Project:

Efforts of this project will continue to be directed toward encouraging collaborative projects involving simulation and modeling of biological systems. Other areas of promise include use of the hybrid computer, because of its flexibility for both analog and digital input and output, for development of prototype models of projects intended for eventual implementation on other real-time systems.

Due to increasing utilization of the hybrid computer, consideration will be given to increasing the core memory size so as to accommodate both real-time and batch tasks on a time-shared basis.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.10
1. Computer Systems Laboratory
2. Laboratory Applications
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer System for Biochemical Laboratories of NIAMD

Previous Serial Number: 4.10

Principal Investigator: Marvin Shapiro

Other Investigators: Jeffrey Buzen, Isaac Hantman, Arthur Schultz, Jay Vinton

Cooperating Units: Laboratories of Physical Biology and Molecular Biology, NIAMD

Man Years:

Total: 4
Professional: 4
Other: 0

Project Description:

Objectives:

To develop a real-time data acquisition and analysis computer system to service laboratories in Building 2, NIAMD.

Methods Employed:

Based on a study of the computational needs of a group of scientists in NIAMD (Building 2), a real-time computer system has been ordered and is scheduled for delivery in June, 1969. Work on the project has proceeded in a number of areas, including the development of: (1) software to handle data acquisition in a real-time environment; (2) interfacing necessary to connect laboratory instruments to the computer; and (3) remote operators' consoles to enable communication between users in the laboratories and the computer.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

Implementation of the proposed computer system would have far-reaching implications: (1) Present turn-around time for computer

results in Building 2 would change from days to minutes. (2) Close interaction between scientists and computer would make new methods of data analysis possible. (3) Real-time data collection would give more accurate results than are now obtainable. (4) The on-line system would be far more flexible and accessible than the current batch processing use of the remotely located IBM 360/50 computer.

Proposed Course of Project:

During Phase 1 of the project, two spectrophotometers, a spectropolarimeter and a Raman spectrometer will be connected directly to the computer. The laboratories involved will have remote operators' consoles and storage oscilloscopes. In addition, a small computer, used to control an X-ray diffractometer, will send and receive digital information to and from the central building computer.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.11
1. Computer Systems Laboratory
2. Laboratory Applications
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer Applications to Mass Spectrometer Data

Previous Serial Number: 4.11

Principal Investigator: Eli Gilbert

Other Investigators: Henry Fales, George W. Milne, J. David Baty, Virginia Aandahl, David Songco

Cooperating Units: Laboratory of Molecular Diseases, NHI

Man Years:

Total: 3.3
Professional: 2.6
Other: 0.7

Project Description:

Objectives:

To apply computer science and technology to determine the topological structure of complex molecules from mass spectrometry data.

Methods Employed:

A small real-time laboratory computer will be used to acquire and reduce raw data. Fragmentation hypotheses are displayed by printer and plotter, along with molecular structure conclusions. A large computer, the PDP-10, with list processing capability does the topological manipulations.

Major Findings:

The on-line laboratory system has been installed, and is working satisfactorily. A rudimentary molecule builder working in LISP is available.

Significance to Biomedical Research and the Program of DCRT:

Through use of the computer, a significant increase in the amount of mass spectrometer data analyzed was realized. In addition, some of the computer output, notably the plotter results, provided new insights into the underlying molecular structures.

Proposed Course of Project:

(1) The on-line system will be upgraded to reduce the need for operator intervention in data acquisition. (2) Direct communication links will be established with larger computers on campus. (3) Heuristic programs for molecular structure will be improved. (4) Applications of learning machines to mass spectral data will be explored. (5) Direct computer control of the mass spectrometer will be used to obtain molecular fragmentation information.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.12
1. Computer Systems Laboratory
2. Systems Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Data Management Systems for the Myocardial Infarction Research Unit Program

Previous Serial Number: 4.12

Principal Investigator: Daniel Syed

Other Investigators: Kenneth Kempner, Martin Miller

Cooperating Units: None

Man Years:

Total:	1
Professional:	1
Other:	0

Project Description:

Objectives:

Nine institutions have been funded by the National Heart Institute to study the causes and characteristics of myocardial infarction. This research program, it is hoped, will lead to improved techniques in both the prevention and treatment of this disease. This project is intended to provide guidance in the establishment of individual data management systems to the nine myocardial infarction research units already funded, and to the three additional institutions scheduled to receive funds. A primary goal is to promote compatibility of hardware and software to the maximum extent possible. A secondary objective is the assimilation of "state of the art" techniques in cardiac monitoring in preparation for the design of a computer system for the recovery area of the Clinical Center heart surgical suite.

Methods Employed:

Guidance to individual myocardial infarction research units was continued through the expedient of providing consultation to the National Heart Institute. In addition, a computer system was designed to accommodate a set of medical protocols stipulated by the National Heart Institute to be typical of MIRU requirements.

The proposed system, if procured by several MIRU's, would allow for common system software and for a collaborative effort in interfacing and applications programming areas, thus alleviating the staffing problems confronting most MIRU's.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

This project provides a basis for evaluating the feasibility of attempting collaboration between institutions establishing individual data management systems directed toward a similar goal. It will also provide experience in cardiac monitoring and care. This experience will be invaluable in prosecuting projects within the heart surgical suite in the Clinical Center.

Proposed Course of Project:

Consultation to the National Heart Institute in the area of the establishment of data management systems will continue. Eventually, this consultation will be extended to include guidance in the analysis of data acquired by the various institutions involved.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.13
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Special Purpose Data Processing Systems for NHI and NCI

Previous Serial Number: 4.13

Principal Investigators: David Songco, Margaret Douglas

Other Investigators: None

Cooperating Units: Laboratory of Metabolism, NHI; Laboratory of Technical Development, NHI; Metabolish Branch, NCI

Man Years:

Total: .3
Professional: .3
Other: 0

Project Description:

Objectives:

To analyze laboratory data collection requirements, and design the appropriate system to collect and prepare the data from these laboratories for computer analysis.

Methods Employed:

The Metabolism Branch, NCI, has purchased two paper tape systems for formatting and storage of scintillation data. A program has been written to perform statistical analysis of the data. During this reporting period, no further action has been taken on the two systems recommended to the Laboratories of Metabolism and Technical Development, NHI.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

Prior to the development of this system for NCI, data analysis was done by hand on small calculators. The computer program and paper

tape data collection greatly reduces this tedious task, allowing the researcher more time to extend his investigations.

Proposed Course of Project:

Refinement of programs for the Metabolism Branch, NCI are planned to encompass additional types of scintillation data.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.14
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Combined Event and Time Code Generator and Reader Units

Previous Serial Number: 4.14

Principal Investigator: David Songco

Other Investigators: Perry Plexico

Cooperating Units: None

Man Years:

Total .5
Professional: .25
Other: .25

Project Description:

Objectives:

To provide a compatible method of annotating analog tape for NIH experimenters that is suitable for manual editing or automatic data processing at established or future data processing facilities.

Methods Employed:

A modified IRIG B BCD time code was chosen as a standard code. At real time, the code is generated in one second frames, with a resolution of one millisecond. The BCD code is arranged to provide hours, minutes, seconds, and day of year or event code. The amplitude modulated 1 K Hz. carrier is suitable for either direct or FM recording, and playback speeds can range from 1/32 to 32 times realtime. The associated CDC 3100 computer interface allows complete control of the generator/reader by software for generating codes, reading codes, high speed analog tape search, or automatic processing of data.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

The availability of a precision time code signal during the recording of experimental data will enable the experimenter to reconstitute time along with his data during later processing of the recorded experiment. A continuous code will greatly facilitate the tedious pre-editing task now employed. In addition, each user will now be using a code that is compatible with any facility with a standard commercially available IRIG B code reader.

Proposed Course of Project:

One computer controlled unit has been completed, and will be interfaced to the hybrid computer in Building 10 in July, 1969. Four manual units have been procured, and will be delivered to various experimenters upon completion of the interfacing work. Future users may purchase standard units from any of several manufacturers.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.15
1. Computer Systems Laboratory
2. Systems Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Beat-to-Beat ECG Analysis

Previous Serial Number: 4.15

Principal Investigator: Daniel Syed

Other Investigators: Kenneth Kempner, Martin Miller, William Holsinger

Cooperating Units: None

Man Years:

Total: 1.5
Professional: 1.5
Other: 0

Project Description:

Objectives:

The primary objective of this research program is the development of real-time ECG rhythm analysis techniques for the detection and classification of cardiac arrhythmias. Milestones in this program are the identification of parameters which will differentiate between the most commonly occurring arrhythmias and the development of computer software and hardware systems for the acquisition and reduction of the ECG waveform.

Methods Employed:

Computer software has been developed to allow analog-to-digital conversion of electrocardiograms and simultaneous analysis for the detection of arrhythmias on a beat-to-beat basis in real-time. Previously developed subroutines detect the presence of R waves, measure QRS width, R and S wave amplitudes, area under the QRS, R-R interval, and the correlation coefficient between the present beat and a standard beat. A diagnosis of current rhythm status is made on the basis of these parameters. ECG's are transmitted to the computer site as a frequency modulated signal on a standard telephone line. A monitoring terminal has been designed and is under construction. This terminal will provide for the acquisition

of ECG's, the control of the computer monitoring process and the display of current rhythm status as determined by the computer.

As a second phase of this project, a small special purpose digital processor has been constructed which, through the use of hardware algorithms, performs an analysis similar to that accomplished via software within the general-purpose digital computer. This device amplified the ECG, detects R waves and determines whether the QRS complex is of normal morphology. On the basis of several consecutive R-R intervals, the rhythm is characterized.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

This project is applicable to the long term monitoring problem in intensive care units and directly supports the MIRU program.

Proposed Course of Project:

Following fabrication and checkout of the remote monitoring terminal, it is hoped that the entire system will be utilized in an extensive series of monitoring runs, with patients in a coronary care environment. During these trials, the software criteria for arrhythmia analysis may be modified, with the aid of cardiologists, to improve the accuracy and scope of the system.

The hardware arrhythmia analyzer will be modified to reflect improvements which were suggested during the initial phases of this project. A second version of this device is also planned and will provide more detailed descriptors of diagnosed arrhythmias.

Honors and Awards: None

Publications:

Kempner, K. M., Miller, M. H., and Holsinger, W. P., "A Computer Approach to Arrhythmia Monitoring," accepted for presentation at the 22nd Annual Conference on Engineering in Medicine and Biology, Chicago, Illinois, July 1969.

Serial No. DCRT - 4.16
1. Computer Systems Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Principles of Global Organization in Dense Volumetric Memories

Previous Serial Number: 4.16

Principal Investigator: Harry Blum

Other Investigators: None

Man Years:

Total:
Professional:
Other:

Project Description:

Project transferred during this reporting period to Laboratory of Applied Studies.

Serial No. DCRT - 4.17
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Data Compaction

Previous Serial Number: 4.17

Principal Investigator: Paul Heffner

Other Investigators: None

Cooperating Units: None

Man Years:

Total: 0
Professional: 0
Other: 0

Project Description:

Objectives:

The objectives of the data compaction project are to reduce memory requirements when voluminous data require storage, and to enable data that would otherwise overload input/output channels to be transmitted and received through the channels.

Methods Employed:

Study has thus far centered around compression techniques that can be applied to continuously occurring analog data that is sampled periodically and stored for later reconstruction and processing. The techniques involved are those which could be implemented either with a computer, or, if appropriate, with external hardware.

Major Findings:

General methods have been derived for data compression; however, their implementation depends upon the specific data to which they are applied--the statistical nature of the signal, the sampling rate, and the margin of error allowed by the experimenter. No work has been done on this project during this reporting period since further work would only be useful on a specific application.

Significance to Biomedical Research and the Program of DCRT:

One of the limiting factors facing on-line data acquisition systems is storage capacity for continuous high data volume experiments. Implementation of the above generated algorithms, either by software or by hardware, will allow continuous storage of data into one storage medium without causing data discontinuity. Some biological data requires fast sampling rates because it is very active and rapidly changing part of the time, even though it is inactive the rest of the time. If this data were to be captured and analyzed in a time-shared computer system, best use could be made of the system if the channel through which that data flows were allowed to be inactive when the data were inactive. A hardware data compressor, external to the computer, would allow for this type of system optimization.

Proposed Course of Project:

When an appropriate application arises, we expect to employ these techniques.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.18
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: File Structure for Medical Information

Previous Serial Number: 4.18

Principal Investigator: William C. White

Other Investigators: Paul R. Moyer, Thomas L. Tubbesing

Cooperating Units: Experimental Virus Isolation Facility, NCI
and Employee Health Unit

Man Years:

Total: .75
Professional: .75
Other: 0

Project Description:

Objectives:

The organization of medical data (health histories, physical examinations, laboratory tests, etc.) into a useful computer-based file is the primary objective of this project. A secondary objective is to satisfy the requirements of the Experimental Virus Isolation Facility (EVIF) program for employee health records.

Methods Employed:

Medical history and physical examination forms were designed, with the assistance of the medical personnel of the Employee Health Unit who were responsible for collecting the data and proofing the key-punched records. The records are being organized in a chronological sequential file, and inverted on all significant words to facilitate retrieval and linguistic analysis of the data.

Major Findings: Not applicable at this time

Significance to Biomedical Research and the Program of DCRT:

This file structure is similar to the organization of the Current Medical Terminology (CMT) file described in another project report

(see 4.7). The experience with these types of records can be useful in the design of hospital information systems and other medical information systems.

Proposed Course of Project:

It is proposed that the EVIF file be turned over to the Data Management Branch for completion and maintenance. Experimental work will continue on methods for the organization and analysis of computer stored medical information.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.19
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Information Retrieval and Natural Language Data Processing

Previous Serial Number: 4.19

Principal Investigator: Milos Pacak

Other Investigators: George Dunham, Helen DeFrancesco, Martin Epstein

Cooperating Units: None

Man Years:

Total: 3.5
Professional: 3.5
Other: 0

Project Description:

Objectives:

The major objectives of the project are to develop the methodology for machine encoding of diagnostic statements into a file, and to provide the capability to retrieve relevant information from the data file with a high degree of completeness.

Methods Employed:

At present, work is proceeding on the natural language input component which will tie into the retrieval system. The system includes the acquisition of textual information, the construction and interrogation of dictionaries, and the necessary grammar and logic to allow for the identification of the information content of the input messages which are the diagnostic statements. These three components must interact to insure that the meaning of the input messages is preserved and captured in a data structure suitable for information storage and retrieval. The messages, that is, diagnostic statements, are not complete English sentences, but are pseudo-sentences written in an outline form, free of verbs and with the noun phrases as the essential syntactic units. To break up the messages, which consist of word forms, into their constituents, algorithms have been developed to determine affixes, to assign grammatical classes to

terms, and to implement transformational rules to derive information content indicators from the text. After a match has been obtained in the dictionary, it is necessary to synthesize an appropriate data structure which can be stored for subsequent retrieval.

Major Findings:

The algorithms developed, thus far, have been related, primarily, to analysis of the input text. As such, they have been successfully employed in encoding over fifty percent of the relatively uncomplicated syntactic structures in the surgical pathology test data.

Expansion of the system requires the development of additional procedures for the recognition of phrase boundaries and for the synthesis of the syntactic and semantic components of diagnostic messages.

Significance to Biomedical Research and the Program of DCRT:

The system will eliminate the need for the hand coding of diagnosis. Data which cannot be recognized as being meaningful will be signaled and returned to the physician for clarification. The linguistic analysis and the development of sophisticated programming techniques employed in the development and implementation of this system are significant to the field of information and computer sciences.

Proposed Course of Project:

1. To improve the structure of currently available medical dictionaries, as necessary, to automatically process the input text.

2. To design a formal semantic model for automated processing of medical English and for the automatic construction of medical thesauri.

3. To expand and generalize the current procedures for automated encoding of pathology diagnosis so that they can be applied in a broader medical context, that is, to clinical medicine.

4. To explore the usefulness of the analytic procedures developed for automatic encoding of medical diagnoses to automatically index and abstract medical documents.

Honors and Awards: None

Publications:

Pratt, A. and Pacak, M., "Identification and Transformation of Terminal Morphemes in Medical English," Methods of Information in Medicine, F. K. Schattauer Verlag, Stuttgart, Vol. 8, No. 2, April 1969.

Serial No. DCRT - 4.20
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Experimental Medical Telecommunications

Previous Serial Number: 4.20

Principal Investigators: William C. White and Scott I. Allen

Other Investigators: Victor Colburn, Michael Otten, Perry Plexico, Eric Swarthe

Cooperating Units: None

Man Years:

Total: 2.5
Professional: 2.5
Other: 0

Project Description:

Objectives:

This project concerns the development and implementation of telecommunication techniques to provide rapid and convenient access to computer based medical information, and the computational capability of large-scale time-shared computer facilities. The goal is to provide the medical profession with low-cost, easy-to-use computer terminals for use from home or office.

Methods Employed:

The general method used in this investigation is to provide medical applications programs, stored on several time-shared computer systems, for demonstration and testing with various input/output terminals. Standard communication terminals, such as the teletype and cathode ray tube display, were investigated, in addition to such non-standard terminals as the "Touch-Tone" telephone, the "Telecopier" facsimile device, and several analog plotters. The pushbutton (Touch-Tone) telephone can be used as a computer terminal for digital input via the tone coded keys and as a receiver for computer driven voice output. Commercial voice answer-back systems were studied, and a speech digitizing and synthesis system is being developed on the hybrid

computer at NIH. Application programs have been written for commercial voice answerback systems, and also modified for use with a small convenient CRT display unit. Other demonstration programs were written to explore the capabilities of the CRT terminal for medical applications. Methods for supplying hard copy computer output to support voice answerback and CRT display output are being investigated. A communication control computer can transmit output directly to the user by using automatic dialing units and facsimile terminals, teletypes, or through telegraph services. The communication control computer is necessary to provide for message buffering and input queuing, code translation and communication with large time-shared computers, as well as supplying the analog controls, relays, and sense registers to operate the telephone data sets and the voice response.

Major Findings:

Commercial voice response systems have proved to be generally unreliable and not adequate for medical telecommunication. The limited vocabulary of the analog systems impose severe constraints on the use of medical terminology. Digitized voice appears to be the method of choice if adequate data compression techniques can be used to minimize the storage requirements. Several vocoder compression methods are being investigated. For more extensive computer output, and to provide for man/machine interaction, the CRT display terminal can be far superior to the conventional teletype.

Significance to Biomedical Research and the Program of DCRT:

The technology being developed can be used to make medical and research information readily available to the research scientists and medical practitioners. The ease of use, low-cost and ready acceptability of telephone input and audio response are receiving consideration in future plans for the main computational facility at NIH as provided by DCRT. The programming techniques and technical requirements for time-sharing, message switching, and analog signal processing for communication by voice grade telephone lines can be of use in future expansion of the system facilities of DCRT and other university medical centers.

Proposed Course of Project:

A Communication Control Computer System will be acquired to implement telephone technology with voice response at DCRT. Continued work on medical applications programs will enable CSL to evaluate the utility and acceptance of the project objectives.

Honors and Awards: None

Publications:

White, W. C., Allen, S. I., Otten, M. and Swarthe, E., "An Experimental Computer Network for Medical Data Processing," Methods of Information in Medicine (in press).

Serial No. DCRT - 4.21
1. Computer Systems Laboratory
2. Laboratory Applications
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Analysis of Laboratory Spectral Data

Previous Serial Number: 4.21

Principal Investigator: Marvin Shapiro

Other Investigators: Karen Bonwit, Marie Chang, Carl Merril, H. Todd Miles

Cooperating Units: Laboratory of Neurochemistry, NIMH;
Laboratory of Molecular Biology, NIAMD

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

To provide a computer analysis of IR and UV spectra and chromatographic data.

Methods Employed:

Mathematical curve fitting techniques and methods of detecting shapes of curves.

Major Findings:

A basic set of programs for analyzing infrared (IR) spectra have been completed and are in frequent use. Programs for analysis of ultraviolet (UV) RNA spectra and for deconvoluting a series of overlapping peaks arising in chromatograph data have been written.

Significance to Biomedical Research and the Program of DCRT:

The computer programs described should eventually have wide applicability at NIH, both by the many laboratories employing the same instruments for which the present programs were written, and by laboratories with other similar types of data which need deconvoluting.

Proposed Course of Project:

Development of the RNA and chromatograph programs is continuing, using real laboratory data.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.22
1. Computer Systems Laboratory
2. Medical Information Science
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Bone Mineral Mass Measurement in-vivo

Previous Serial Number: 4.22

Principal Investigator: William C. White

Other Investigators: William Vincent

Cooperating Units: Clinical Endocrinology Branch, NHI

Man Years:

Total: 0
Professional: 0
Other: 0

Project Description:

This project has been completed, and turned over for production to the Data Management Branch, DCRT.

Serial No. DCRT - 4.23
1. Computer Systems Laboratory
2. Systems Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Remote and Satellite Computation for National Institute of Mental Health Building 10 Scientists

Previous Serial Number: 4.23

Principal Investigators: Daniel Syed, Victor Colburn

Other Investigators: William Holsinger, Martin Miller

Cooperating Units: None

Man Years:

Total: 0.5
Professional: 0.5
Other: 0

Project Description:

Objectives:

This project is intended to provide both on-line real-time and off-line computational support for research programs of National Institute of Mental Health scientists who are currently scheduled to remain in Building 10. Specifically, on-line real-time data acquisition and computation is proposed in support of experiments in Problem-Solving, Learning, Sleep and Dreaming, and Perception, and also in support of experiments relating the analysis of electroencephalographs and physiological variables to clinical diagnosis and classification of NIMH patients. Off-line computation is proposed in support of studies involving voluminous statistical processing as in the analysis of data from various surveys.

Methods Employed:

The display requirements of Building 10 NIMH scientists were restudied. Modifications to the originally proposed on-line real-time system resulted in the recommendation of a dual central processor system equipped with a versatile display capability. Manufacturers' proposals have been received and are currently under review. Off-line computational requirements have been met by the installation of a remote IBM 360/20 terminal, operating under "remote job entry" mode to the central facility's IBM equipment.

Major Findings:

The feasibility and utility of providing a remote queued batch processing capability to institutes around the campus was established. Improvement of this capability will be effected by the Central Computer Branch and will not constitute a goal of this project.

Significance to Biomedical Research and the Program of DCRT:

Consistent with the development of time-sharing software on the Central Computer Facility IBM 360/50, the system will provide a basis for analysis and development of satellite computing techniques in a biomedical environment characterized by the presence of both high volume off-line data processing requirements and sophisticated on-line real-time requirements.

Proposed Course of Project:

Upon completion of the on-line computer system procurement phase, system software and applications programs will be developed. Interfaces and investigator response panels will be designed and fabricated for the experiments to be automated. A computer-to-computer link with the Central Computer Facility will probably be implemented.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.24
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Gerontology Research Center Computing System

Previous Serial Number: DCRT 4.24

Principal Investigator: Perry S. Plexico

Other Investigator: Isaac Hantman

Cooperating Units: None

Man Years:

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

This project is intended to provide support for research programs of the Gerontology Research Center, Baltimore, in the areas of on-line computation and control, analog-to-digital conversion, and off-line data processing. The on-line computation capability is designed to support experiments in problem-solving, while the on-line control capability is to be utilized in the control of blood glucose/insulin infusion rates. The analog-to-digital conversion and off-line computation capabilities will be used in the analysis of electrocardiograms, physiological data, scintillation counter data, and nuclear magnetic resonance data, etc.

Methods Employed:

The analysis of data processing requirements and system specifications were completed and submitted through departmental channels to obtain data processing clearance. Program approval was received in March, at which time a request for proposals was released. Responses to this request are being evaluated.

Major Findings: None

Significance to Biomedical Research and the Program of DCRT:

Implementation of the proposed computer system will provide support to research programs not obtainable from any other source, particularly in view of the remote location of the Gerontology Research Center with respect to the NIH Campus.

Proposed Course of Project:

Upon selection of a specific system and negotiation of an acceptable contract, hardware for interfacing real-time experiments to the computer will be developed. Support for software generation to permit efficient use of the computer system by concurrent execution of both real-time and off-line tasks will be provided to the extent allowed by available resources.

Honors and Awards: None

Publications: None

Serial No. DCRT - 4.25
1. Computer Systems Laboratory
2. Systems Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer System for the Laboratory of Psychology, Section on Comparative Behavior, Poolesville

Previous Serial Number: 4.25

Principal Investigator: Daniel Syed

Other Investigators: Martin Miller, Kenneth Kempner

Cooperating Units: None

Man Years:

Total: 0
Professional: 0
Other: 0

Project Description:

Objectives:

This project was designed to provide a real-time data acquisition computer system in support of on-line experiments being conducted by the Section on Comparative Behavior, Laboratory of Psychology, NIMH. Because of the difficulty in obtaining funds, and because of the unexpected availability of a surplus compromise system, the project was terminated.

Serial No. DCRT - 4.26
1. Computer Systems Laboratory
2. Processor Design
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Wave Propagation in Arteries

Previous Serial Number: None

Principal Investigators: Perry S. Plexico, Ivor Gabe

Other Investigators: Donald L. Fry, Dali J. Patel

Cooperating Units: Section on Clinical Biophysics, CB, NHI

Man Years:

Total:	1
Professional:	1
Other:	0

Project Description:

Objectives:

1. To apply and test the validity of an analogy between arterial pressure wave transmission and transmission via electrical lines.

2. To quantitate the parameters of such a model in terms of physical characteristics of the vessel.

Methods Employed:

Simultaneous pressure recordings were taken at three points along the descending thoracic aorta of 14 dogs. These were converted from analog-to-digital form using the hybrid computer, and then subjected to Fourier transformation. The transformed data was then used in evaluating solutions to the telegraph equations in order to compute the arterial propagation constant for each of the 14 dogs studied. This quantity is a complex function of frequency involving the pressure wave attenuation and velocity which are, in turn, related to physical characteristics of the artery, such as its elasticity.

Major Findings:

Comparison of arterial wave velocity calculations, as determined above, with results obtained by other methods (e.g., from direct

measurement of vessel elasticity, and with Wormersley's predictions) showed a good correlation. Wave attenuation results, on the other hand, could not be adequately explained by a simple unifrom transmission line model. A more complex analogy, that of an exponentially tapered transmission line, was somewhat more satisfactory; however, the degree of taper predicted by the model substantially exceeded that which could be experimentally verified. Possible conclusions to explain these findings include (a) the existence of real but unmeasurable hydraulic tape (as opposed to geometric tape), (b) the existence of tape which is not exponential in nature, and (c) the existence of extreme non-linearities in the properties of the artery.

Significance to Biomedical Research and the Program of DCRT:

This work has broad potential, as does other work relating to development of models of arterial flow, in achieving a better understanding of the cardiovascular system, and specifically, the disease processes which affect it.

Proposed Course of Project:

After completion of analysis of all results, it is expected that findings will be submitted for consideration for publication in. Circulation Research.

Honors and Awards: None

Publications: None

July 1, 1968, through June 30, 1969

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH

DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities | 1. DCRT - 5

2. PHYSICAL SCIENCES LABORATORY | 3. Dr. G. H. Weiss

I. SUMMARY

The Physical Sciences Laboratory has been staffed with scientists whose skills lie in various areas of physics, mathematics, and chemistry. The Laboratory thus functions to enrich the N.I.H. community with a broad capability in mathematical physics and theoretical chemistry. Individual staff members have been chosen because of their personal interest in performing research on problems related to both the biomedical and physical sciences.

Included among the research projects undertaken during the past year are theoretical and experimental studies of the conformations of biologically interesting molecules, the development of models and theories to explain various membrane phenomena, theoretical studies of macromolecular association processes and chemical kinetics, theoretical studies related to the application of physical instruments to the study of biological molecules, and the development of computer assist devices for the analysis of data from biomedical experiments. The number of professionals on the Laboratory staff has averaged eight during FY 1969.

II. CURRENT BRANCH PROGRAMS

A. Objectives

To develop a group with theoretical interests to analyze biological phenomena in the terminology of theoretical physics and theoretical chemistry. To act as consultants in theoretical aspects of physics and chemistry to experimental scientists at NIH and to members of DCRT who may require such services in conjunction with their own work.

B. Progress of Current Programs

1. Development of Ultracentrifugation Theory

a. Theories have been developed for various techniques of centrifugation of molecules of intermediate size.

b. The theory of separation of tumor cells in a Ficoll gradient has been perfected.

2. <u>Theory of the Helix-Coil Transition in Polypeptides</u>

a. Theories of the helix-coil transition have been modified to account for recent experimental results. Further experiments on this phenomenon are being carried out.

3. <u>Computer Processing of NMR Data</u>

a. Numerical procedures and computer programs have been developed for the interpretation and Fourier analysis of NMR data, particularly for the resolution of spectra. These programs are presently being used by members of NIAMD.

4. <u>Molecular Model Building</u>

a. Two general methods for deriving molecular interaction energy functions from equilibrium gas phase and crystal phase data have been developed. Solutions for the accurate evaluation of conditionally convergent terms in infinite lattice sume of the R^{-1} type have been discovered.

5. <u>Biophysical Analysis</u>

a. A model has been developed to explain fluctuations in the firing threshold of isolated nerve axons. The model predicts the shape of the threshold firing curves as a function of various experimental parameters.

b. Solutions have been obtained to the kinetic equations which characterize schemes pertaining to formation of macromolecular complexes between cationic polypeptides and DNA.

c. Various properties of excitable transport units obtained from aerobacter cloacae have been studied; theories of cooperativity in biological membranes have been formulated.

6. <u>Configurational Statistics of Polymer Chains</u>

a. Theoretical Studies have been initiated which are directed towards determining the configurations of linear polyelectrolyte chains.

7. <u>Excitation and Transport Properties of Fluids</u>

a. A previously derived theoretical description of the collective excitations of fluids has been extended to include phonon lifetimes.

b. Calculations have been performed related to laser scattering from solutions of biological macromolecules.

8. Theory of Cell Membranes

a. Theoretical studies show that membranes may deform under the influence of traversing charged ions.

b. Calculation of van der Waals dispersion forces in the regime of a thin lipid film suggest that these are important contributions to interfacial energy, but not as a long-range force across the membrane.

c. Electrostatic analysis shows that pH and ion concentration in the microenvironment of the charged cell membrane can differ considerably from that of the extra cellular space. It is likely that the activity of membrane enzymes depends on this microenvironment.

9. Consulting Services

a. Computer programs have been developed to fit experimental NMR spectra of intermolecular AB exchanging systems.

b. Mathematical descriptions of protein systhesis in mammalian cells have been formulated.

c. Numerical techniques have been perfected to correlate data from circular dichroism and ORD experiments.

10. Fundamental Studies

a. Calculations of the number of leads required for accurate determination of the equivalent heart dipole have been completed.

b. Various models of enzyme systems have been analyzed.

c. Calculations have been performed to provide current-voltage curves for model membranes.

11. Modelaide

a. A computer assist device has been designed enabling a research scientist to simultaneously present data and mathematical curves on a graphics display terminal.

C. Program Plans

Research activities similar to those described above will be performed. Emphasis will be given to obtaining better understanding of the properties and functions of biological macromolecules and cell membranes. Consultantship services and basic research in chemical physics will be sustained; computer studies of macromolecular configurations will be continued.

It is anticipated that somewhat greater emphasis will be given to development of lasers as physical probes of systems of biological molecules. If proposed personnel actions are consummated, additional projects will be undertaken related to NMR studies of biologically interesting molecules.

Serial No. 5.1
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Theory of the Ultracentrifuge

Previous Serial Number: 5.1

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: Ralph Nossal, Ph.D., Richard I. Shrager

Cooperating Units: Charles W. Boone, M.D., NCI

Man Years

Total:	.50
Professional:	.40
Other	.10

Project Description:

Objectives:

To determine the effects of various factors such as concentration dependent sedimentation, pressure, density gradients, variations in rotor speed, and polydispersity on current techniques for determining molecular weights. To devise corrections and new techniques of ultracentrifugation which bypass or eliminate these effects.

Methods:

The methods employed include numerical solutions to partial differential equations and classical analysis.

Major Findings:

Formulae have been developed for the early sedimentation of molecules of intermediate size. Procedures have been developed to separate tumor cells in a Ficoll gradient. The theory of band centrifugation in non-ideal solutions has been developed.

Significance to Biomedical Research

Density gradient centrifugation is one of the most significant modern tools for separating molecular species and for populations of macroscopic tumor cells. Band centrifugation is of considerable importance in the study of biologically interesting molecules since very little material is required for such experiments.

Honors and Awards: None

Publications:

Nossal, R.J., and Weiss G.H.: Early time expansions for sedimentation of weakly non-ideal solutions. Biopolymers (in press).

Weiss, G.H., Pretlow, T.G., Boone, C.W., and Shrager, R.I.: Rate zonal centrifugation in a ficoll gradient. Analytical Biochemistry. 29: No. 2, 230-237, 1969.

Serial No. 5.2
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Theory of the Helix-Coil Transformation of Polypeptides in Solution

Previous Serial Number: 5.2

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: V. A. Parsegian, Ph.D., B. W. Ninham, Ph.D.

Cooperating Units: Laboratory of Organic Chemistry, NHI

Man Years:

Total:	.5
Professional:	.5
Other	0

Project Description:

Objectives:

To determine the effects of various factors on the stability of helical polypeptides and to formulate models for the helix-random coil transformation by the application of statistical thermodynamics.

Methods:

Applying NMR spectroscopy as the experimental approach towards evaluating the relative importance of the various factors and then formulating models which are consistent with the experimental data.

Major Findings:

The helix-random coil interconversion rate is slower than 0.1 sec^{-1} which suggests that previous models do not adequately describe the phenomenon. Poly-γ-benzyl-L-glutamate shows anomolous behavior above molecular weight 40,000 which suggests that a three-dimensional "whole molecule" picture is necessary to describe the transformation.

Significance to Biomedical Research

Polypeptides serve as protein model compounds. The observation of a three-dimensional effect suggests that such a model may provide a mechanism for the storage of energy in enzymatic proteins. Such a picture is necessary to understand both binding properties of enzymes and the phenomenon of protein denaturation.

Honors and Awards: Election as a fellow of the American Institute of Chemists.

Publications:

Ferretti, J.A., and Paolillo, L.: An NMR investigation of the melix to random coil transformation in poly-α- amino acids. I. poly-L-alanine. Biopolymers. 7: 155-171, 1969.

Feretti, J.A.: NMR studies of the helix-random coil transformation in poly-α- amino acids. Polymer Preprints. 10: 29-35, 1969.

Serial No. 5.3
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computer Processing of Nuclear Magnetic Resonance (NMR) Spectral Data

Previous Serial Number: 5.3

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Mildred McNeel, Edwin D. Becker, Ph.D. (LPB:NIAMD), Rolf Johannasen, Ph.D. (NBS), T. C. Farrar, Ph.D. (NBS) and S. Druck (Catholic University).

Cooperating Units: NHI, NIAMD

Man Years:

Total:	1.2
Professional	1.1
Other	.1

Project Description:

Objectives:

To develop methods to analyze complex frequency domain and time domain (impulse response) NMR spectra and to Fourier transform impulse response spectra to frequency domain spectra. To devise methods for computer simulation of experimental spectra and to decompose these spectra into their individual component frequencies and intensities.

Methods:

An iterative computer program has been written to approximate frequency domain experimental NMR spectra using a least squares technique. A fast Fourier transform program is employed to obtain frequency domain spectra from the impulse response spectra.

Major Findings:

It is possible to completely analyze complex NMR spectra with the use of a high speed computer. The Fourier transform method has been

applied to obtain carbon-13 spectra in natural abundance of a number of molecules. The results show that, using the Fourier transform technique a 100-fold enhancement of carbon-13 sensitivity for a given observation time is possible. Alternatively, for a given signal-to-noise ratio, the observation time will be reduced by a factor of 1000 or more when compared with conventional scanning techniques.

Significance to Biomedical Research:

Carbon-13 NMR provide information about the conformations of macromolecules and about the active binding sites of enzymes.

Honors and Awards: None

Publications:

Ferretti, J. A., Johannasen R., and Harris, R. K.: Analysis of the NMR spectrum of Triisopropylphosphine. A new iterative computer program. Journal of Chemical Physics (in press).

Serial No. 5.4
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Molecular Model Building Using Theoretically and Empirically Determined Intra- and Intermolecular Potential Functions

Previous Serial Number: 5.4

Principal Investigator: William P. Minicozzi, M.D.

Other Investigators: Michael Stroot, James Kiefer.

Cooperating Units: None

Man Years:

Total:	2.5
Professional:	2.0
Other:	0.5

Project Description:

Objectives:

To make possible the understanding and solution of biological problems on a detailed molecular basis by developing models which accurately account for the force fields that molecules experience.

Methods Employed:

Theoretical criteria for deriving molecular interaction energy functions from equilibrium gas and crystal phase data are determined by using principles of physics. Experimental data are modified to be consistent with the theoretical criteria. Mathematical models are constructed, tested, and, where necessary, altered so as to be in accordance with the above mentioned criteria. Mathematical methods to make the above procedure feasible are developed as needed. Finally, programs are implemented to perform the above calculations on the 360/50 computer.

Major Findings:

Two general methods have been developed for deriving molecular interaction energy functions from equilibrium gas phase and crystal phase data.

Solution for the accurate evaluation of conditionally convergent terms in infinite lattice sums of the R^{-1} type have been discovered.

Significance to Biomedical Research

The successful development of this approach would make it possible to define and solve biological problems on a detailed molecular basis. This approach can probe in areas presently unaccessable by other means.

Proposed Course of Study:

To generalize this approach so as to make it applicable to all phases of a physical system.

Honors and Awards: None

Publications:

Minicozzi, W.P., and Bradley, D.F.: On the determination of interaction energy surfaces. I. Formic and acetic acid dimers. J. Comp. Physics. 4, 555-575, 1969.

Serial No. 5.5
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Biophysical Analysis

Previous Serial Numbers: 5.5, 5.6

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: Barry W. Ninham, Ph.D, J. A. Ferretti, Ph.D.

Cooperating Units: Harold Lecar, Ph.D., Gerald Ehrenstein, Ph.D.
Biophysics Laboratory, NINDB

Man Years:

Total:	1.05
Professional:	.95
Other	.10

Objectives:

To supply theoretical foundations for various experimental observations which arise in physiology and biophysical chemistry. The following problems have received particular attention during the past year:

1) Threshold fluctuations in nerves: Objectives are to relate the fluctuations in firing thresholds of nerves to the chemical and physical processes underlying excitation and to analyze existing data in order to test theoreis concerning mechanisms of transport of ions across nerve membranes;

2) Associations between cationic polypeptides and DNA: Objective is to provide theories to distinguish between various possible mechanisms by which polyamines and polypeptides bind to DNA;

3) Cooperative phenomena occuring in phospholipid bilayer membranes: Objective is to develop experiments and theories relating conductivity measurements and switching noise data to the molecular processes which give rise to excitability.

Methods:

1) Existing neurophysiological equations (the Hodgkin-Huxley equations) have been modified to include fluctuating forces. Methods of non-linear mechanics and statistical physics have been applied to analyze the equations.

2) Kinetic equations have been derived and investigated for a number of possible association mechanisms.

3. Various techniques of mathematical physics have been modified and applied to describe the current noise which appears when constant voltages are applied across excitable biological membranes and their analogs.

Major Findings:

1) The previous results, whereby the functional shape of the threshold firing curve and the width of the threshold region were related to models of internal noise and conductivity fluctuations, have now been generalized to include other noise sources. Higher order effects due to non-linearities have explicitly been investigated.

2) Results have been obtained for the equilibrium solutions and approach to equilibrium of various sets of kinetic equations.

3) Experimental procedures have been developed to voltage clamp excitable transport units obtained from aerobacter cloacae. Single units, having a conductivity of $\sim 3 \times 10^{-10}$ mho, have been isolated and their properties studied. Also, experiments have been performed with membranes which contain many such units, and a theory has been devised to discern whether the units interact with each other.

Significance to Biomedical Research:

1) Threshold fluctuations: It is hoped that the study will facilitate better understanding of the physical processes underlying the excitation and propagation of nervous impulse.

2) Polypeptide binding to DNA: The theoretical studies describe systems which are analogs of various physiological phenomena, among them being the association of histones with the nuclei of cells.

3) Reconstituted phospholipid membranes: The membrane components can be controlled by the experimentalist, thus providing an opportunity to devise procedures and test theories applicable to natural biological membranes.

Publications: None

Serial No. 5.6
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Configurational Statistics of Polymer Chains

Previous Serical Number: None

Principal Investigators: B. W. Ninham, Ph.D., I. Darvey, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years:

Total:	.85
Professional:	.75
Other:	.10

Project Description:

Objectives: Theoretical analysis of the asymptotic solution of the excluded volume problem in a linear polymer chain, with particular reference to polyelectrolytes.

Methods:

Mayer cluster-expansion and Bogoliubov-Born-Green-Kirkwood-Yvon hierarchy equations of statistical mechanics.

Major Findings:

No Major results have yet been obtained. The problem has only recently been undertaken.

Significance to Biomedical Research:

The statistical mechanics of ordinary polymers when excluded volume effects are taken into account is one of two possible starting points - as opposed to direct computation - for investigations on the configurational statistics of biological macromolecules. Many biopolymers are polyelectrolytes, and previous theories have been on an ad hoc basis. A first principle treatment of the configurational statistics of polyelectrolytes as a function of salt concentration should be of importance to an understanding of biological mechanisms on a molecular level.

Serial No. 5.7
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Excitation and Transport Properties of Fluids

Previous Serial Number: 5.7

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years:

Total:	.20
Professional:	.20
Other:	.0

Project Description:

Objectives:

To provide basic knowledge concerning the excitation properties and transport properties of both simple fluids and complex solutions containing biological macromolecules.

Methods:

Theoretical techniques of mathematical physics and statistical mechanics are employed in order to develop new physical theories. Theoretical studies are performed in support of new experiments, particularly those invovling the scattering of light from laser sources.

Major Findings:

A previously derived theoretical description of the collective excitations of fluids has been extended to include phonon lifetimes. Various calculations have been performed related to laser scattering from solutions of biological macromolecules and a series of experiments have been planned. Some ancillary calculations of the motion of molecules in the analytical ultracentrifuge have also been performed.

Significance to Biomedical Research:

Almost all biological phenomena occur in a fluid environment. A number of fundamental questions concerning the physical behavior of fluids yet remain unanswered. Their elucidation will ultimately facilitate better understanding of the functions and properties of biological systems.

Publications: None.

Serial No. 5.8
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 20, 1969

Project Title: Intermolecular Forces Acting in the Cell Membrane

Previous Serial Number: 5.8

Principal Investigator: V. Adrian Parsegian, Ph.D.

Other Investigators: B. W. Ninham, Ph.D.

Cooperating Units: None

Man Years:

Total:	1.35
Professional:	1.25
Other:	.10

Project Description:

Objectives:

To identify and calculate those intermolecular forces important in stabilizing the cell membrane and in determining transport of materials across the cell boundary. These have been coulombic interactions between charged species (expecially between the cell wall and its ionic environment), dispersion (van der Waals) forces acting both at an interface and in cell-cell interactions and electrostatic forces between a molecule and its polarizable environment.

Methods:

Classical and quantum-mechanical treatment of electromagnetic and statistical-mechanical behavior as well as molecular models of specific molecular associations.

Major Findings:

Ionic interaction with a dielectric unit cell membrane is strong enough to exert a deformative force on the membrane it traverses; it is no longer correct to treat the membrane as a pre-existing structure posing a fixed obstacle to permeating charged species. This may have

important implications in phenomena such as nervous excitation where membrane permeability depends on the flow of material across it.

The pH in the microenvironment of an enzyme at a cell wall or cell membrane may be as much as two units lower than the pH of the surrounding medium. Analysis of this phenomenon in terms of the surface density of ionizable groups and specific dissociation constants of surface groups on the cell wall now provides a method for calculating pH and ion concentrations in the vicinity of the cell membrane.

We have successfully applied the very powerful and general method of Lifshitz to calculate the van der Waals force of water across a thin lipid membrane. This is the first application of this method (or any rigorous method) to a problem of biological importance. The theory correctly predicts the experimentally observed strength of the dispersion force and, more important, reveals an unexpected effect of salt concentration on the electromagnetic dispersion forces.

Significance to Biomedical Research:

The electrostrictive force exerted by an ion upon a membrane barrier may explain time varying membrane permeability changes during rapid ion transport in cells.

The activity of enzymes attached to cell surfaces is a function of its microenvironment; changes in microenvironment are a means to control activity. We have a method now of describing that environment in terms of known molecular parameters.

Dispersion forces between cells may exert a force for cell-cell contact and "recognition" (as well as a force determining membrane structure). We are now in a position to evaluate the importance of these interactions using a rigorous and sophisticated physical theory.

Honors and Awards: NIH Suggestion Award

Publications:

Parsegian, V.A.: The energy of an ion crossing a low dielectric membrane: solutions to four relevant electrostatic problems. Nature 221, 844-846, 1969.

Parsegian, V.A.: Dielectric aspects of biological materials. Annual digest of literature in dielectrics (in press).

Serial No. 5.9
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Consulting Services

Previous Serial Number: 5.9

Principal Investigators: George H. Weiss, Ph.D., Mildred McNeel, Richard I. Shrager

Other Investigators: Gitte Shoup, Ph.D., Elliot Charney, Ph.D., Gordon Tompkins, M.D.,Ph.D., Eric Johnson, M.D.

Cooperating Units: Laboratory of Physical Biology, NIAMD, Laboratory of Molecular Biology, NIAMD, Surgery Branch, NCI.

Man Years:

Total:	.60
Professional:	.50
Other:	.10

Project Description:

To provide consulting services in biometry, applied mathematics, theoretical physics and chemistry to workers who are primarily in experimental fields.

Methods:

The methods include statistical analysis, theoretical physics and chemistry, and applied mathematics.

Major Findings:

Computer programs have been developed to fit experimental NMR spectra of intermolecular AB exchanging systems.

Mathematical descriptions of a) protein systhesis in mammalian cells and b) tumor cell breakdown in the presence of spleen cells have been formulated.

Numberical techniques have been perfected to correlate data from

circular dichroism and from optical rotatory dispersion experiments, and to decompose them into spectral components.

Honors and Awards: None

Publications: None.

Serial No. 5.10
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Fundamantal Studies

Previous Serial Number: 5.10

Principal Investigators: George H. Weiss, Ph.D., Barry W. Ninham, Ph.D., Ralph J. Nossal, Ph.D., Gershom Zajicek, Ph.D., Ivan Darvey, Ph.D., Michael Mackey, Ph.D. Richard I. Shrager

Other Investigators: Irwin Oppenheim, Ph.D., Robert Zwanzig, Ph.D., Leonard Kohn, M.D.

Cooperating Units: Laboratory of Biochemical Pharmacology, NIAMD.

Man Years:

Total:	1.2
Professional:	.9
Other:	.3

Project Description:

This project encompasses several lines of investigation, some, but not all of which relate to biomedical problems. These include theoretical studies of red cell kinetics, the calculation of diffusion properties of long chain polymers, the analysis of models of enzyme systems, development of theories of membrane transport, design studies for a new electrocardiograph, and basic studies in statistical mechanics.

Methods:

The methods used include partial differential equations, statistical theory, classical analysis, numerical analysis, and other techniques commonly used in theoretical physics.

Major Findings:

Calculations of the number of leads required for accurate determination of the equivalent heart dipole have been completed. A number greater than 100 will provide errors of less than 15%.

Various models of enzyme systems have been analyzed. One model, for a system of two enzymes in which one of the substrates of the first enzyme is regenerated during the process by the second enzyme, may provide an explanation for inhibition of the myosin-creatine kinase system at high enzyme concentrations.

Calculations have been performed to provide current-voltage curves for model membranes composed of fixed scattering centers which interact with charged current carriers according to various force laws.

Significance to Biomedical Research:

The work on cell kinetics may lead to new interpretations of tissue growth. The work on the electrocardiograph is directed towards development of new techniques in pediatric cardiology. Enzyme kinetics provides information on the mechanisms of biological reactions.

Publications:

Weiss, G.H.: Equations for the age structure of a growing population. Bull. Math. Biophys. 30, 427-435, 1968.

Weiss, G.H., and Fischmann, E.J.: Effect of reducing the number of surface potential measurements on estimates of heart dipole moments. Trans. IEEE Biomed. Engr. (in press).

Serial No. 5.11
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Modelaide: A Computer-Graphics Program for the Evaluation of Mathematical Models.

Previous Serial Number: None

Principal Investigator: Richard I. Shrager

Other Investigators: None

Cooperating Units: Systems Team, particularly William Speary

Man Years:

Total:	1.0
Professional:	0.9
Other:	0.1

Project Description:

Objectives:

The object of Modelaide is to lower the computation barrier between an investigator and his mathematical notions about the workings of an observable phenomenon.

Methods:

The mathematical methods are from two fields; non-linear regression and mathematical programming. The program itself is coded in PL/1, and makes extensive use of character strings, bit strings, and interrupt handling. A ten-hour course is offered by DCRT in the use of Modelaide, and a program manual is available.

Major Findings:

A system has been designed so that program flow may be directed by the investigator from a graphics display terminal. He can request curve fits, input from a prepared tape, output on the display, printer, or calcomp plotter. The models being used, the graphs and the data transformations are limited only by what is available in the programmed repetoire, or by what the investigator can design himself. Once the data is entered, he can alter, add, delete, deactivate, reactivate, print, plot, or display any item in the standard data set. In other words, Modelaide provides access to powerful curve fitting techniques, and responds quickly to many related requests, including graphic display. Several investigators have used the system (some after taking the course), and have found it impressive. Their models range from sums of gaussians to non-linear differential equations.

Significance to Biomedical Research:

Modelling, in which Modelaide is helpful, can serve as verification, simulation, and a guide to future experiments.

Proposed Course of Project:

Use of the program in worthwhile projects is the major goal.

Honors and Awards: None

Publications: None

July 1, 1968 through June 30, 1969

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT 6 (Serial Number)

2. HEURISTICS LABORATORY — 3. James R. Slagle (Chief)

I. SUMMARY

The Heuristics Laboratory now includes eight scientific and two clerical individuals. The new PDP-10 computer has been delivered, passed acceptance tests, and is now fulfilling most of the computing needs of the Laboratory.

Progress on the question-answering project has been good. Several preliminary objectives have been attained and a number of papers have been published, some theoretical and some experimental.

Much of the preliminary software needed for image processing has been completed and processing of actual biomedical photos is now beginning. General theoretical papers have been published.

A new project, Pattern Recognition, has been started. The general pattern recognition problem is of fundamental importance and is basic to both image processing and general problem solving. Several experimental and several theoretical papers have been written.

II. CURRENT LABORATORY PROGRAMS

The Heuristics Laboratory was established in January 1968. It was formed from newly recruited personnel and the Biological Image Processing Group which was transferred from the Computer Systems Laboratory. Effective April 30 of this reporting period, the Biological Image Processing Group was transferred from the Heuristics Laboratory to the Computer Center Branch. During this reporting period, three projects were active in the Laboratory.

A. Objectives

This Laboratory has five broad objectives:

1. To develop concepts, techniques and programs for the automatic answering of questions and the solving of problems.

2. To design and develop a computing system and the associated programming techniques for the on-line, interactive processing of biological images.

3. To maintain a reservoir of expertise in the area of heuristic programming which can be tapped by others for advice and assistance on other computer-related projects.

4. To collaborate with other divisions and laboratories on biomedical projects.

5. To develop techniques in automatic pattern recognition as needed in the above areas.

B. Progress

The objectives above represent the broad goals of the Laboratory. Progress toward each of these objectives is given below. More detail will be found in the individual project reports.

Progress Toward Objective #1

There are many biomedical problem areas in which deductive reasoning from facts must be done. Past studies have shown that computers can do deductive reasoning, however, serious difficulties stand in the way of realistic applications. They are:

1. Question Formulation. A poor formulation can make a question effectively unanswerable.

2. Time Required for Deduction. Prohibitive time may be required for deduction of conclusions from a given set of facts.

3. Selection of Facts. In a large data base, only a few facts will be relevant to a given question. There must be some way to select the appropriate facts.

In order to overcome the first difficulty, studies in the formulation of questions and problems have begun.

The second difficulty indicates a need for improved strategy or control over the direction of deduction so that effort will be focused on the most fruitful lines of reasoning. Faster deduction techniques may also be helpful.

The third difficulty implies a need for high level planning techniques to determine what data should be retrieved from the data base. Both theoretical and practical progress has been made in all of these areas.

Sufficient progress has been made that we can now begin work on a large scale question-answering system which will integrate the various techniques which have been developed.

Progress Toward Objective #2

The PDP-10 computer with special attached image processing hardware has been delivered and has passed acceptance testing. A great deal of image processing software has been written and debugged. Basic techniques needed for on-line interactive image processing have been demonstrated. Actual work with real

biological photos is now beginning. This project has been transferred out of the Heuristics Laboratory effective April 29, 1969. Hence, the next report in this area will be under the Computer Center Branch.

Progress Toward Objective #3

Advice and assistance on several outside projects has been given by the Heuristics Laboratory.

A presentation of heuristic programming techniques was given to the Data Management Branch and other interested personnel.

A series of classes on the LISP computer language was given by Richard Feldmann.

Assistance was given to the Computer Center Branch on the 360 LISP system.

Progress Toward Objective #4

Heuristics Laboratory personnel have collaborated with others on several inter-laboratory projects.

A program to solve problems in organic chemistry was developed in collaboration with Eli Gilbert.

Image processing work has been carried out in collaboration with Dr. Lewis Lipkin.

Progress Toward Objective #5

General theoretical papers on pattern recognition have been written. Two programs which do diagnoses by pattern recognition were written and tested.

General Remarks

The use of the new PDP-10 computer in on-line mode has substantially improved the ability of the Laboratory to carry out computational projects rapidly.

The long-range significance of the heuristic programming field is discussed by Kahn and Wiener on page 89 of "The Year 2000".

Serial No. DCRT 6.1
1. Heuristics Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Automatic Question-Answering

Previous Serial Number: Same

Principal Investigator: James R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D., John K. Dixon, Carl Farrell, Deena Koniver, Richard C.T. Lee, Ph.D.

Cooperating Units: None

Man Years
Total: 7.2
Professional: 5.7
Other: 1.5

Project Description:

Objectives:

1. To develop concepts and formalisms for the representations of facts, questions, and problems.

2. To develop strategies which reduce the amount of computation required to answer questions and solve problems.

3. To develop techniques to extract, from a large data base, those facts which are relevant to a given question.

Methods Employed:

As new concepts arise, internal papers are generated. The computational feasibility of these proposals is studied and computer programs are written and debugged. Most computation is now being done on the new PDP-10 computer. The IBM 360/50 has been used during this year but is now used to a lesser extent. The Q-32 computer in Santa Monica, California is no longer used. Most programs are written in the LISP language.

Major Findings:

Theoretical developments concerning fuzzy sets in euclidean space have been made which may be useful in the formulation of questions and problems.

Great strides have been made in improving the strategy of deduction. Seven procedures have been programmed and debugged. Eight papers have been written. Extensive experiments have demonstrated that the MULTIPLE and the M & N tree searching procedures are more efficient than previous techniques. A learning method has been implemented for the MULTIPLE program and preliminary results indicate success. Improved deduction procedures called Renamable Paramodulation, Input Resolution, Pattern Recognition Theorem Proving, Smallest Pair Resolution, and Learning Paramodulation have been programmed and some testing has been done. Theoretical papers have been written on Pattern Recognition Theorem Proving, and Semantic Resolution. In general, each of these procedures has some advantages over the previous techniques.

Two programs have also been written which automatically write programs. One of these, called PROW, uses deduction techniques. The other, called SPECIALIZER, uses a new principle. Both produce useful programs. It is clear that a highly intelligent and efficient question-answering system will have to write special purpose programs for itself and then execute them. PROW and SPECIALIZER are a step in that direction.

Significance to Biomedical Research and the Program of DCRT:

The biomedical significance of this project has two main aspects:

1. A fully developed question-answering problem-solving program would allow the working scientist to turn over to the computer routine questions and problems. The computer would solve these routine matters while the scientist was attending to the more important and difficult problems. It is expected that the question-answering program will be sufficiently general-purpose, that it will be able to operate in many areas simply by changing the data base and with little or no re-programming.

2. The greatest value of this project from a long term point of view will perhaps be its contribution to fundamental knowledge. Extrapolating from present knowledge, it seems reasonable to believe that when the biochemistry of the human body becomes more completely known, the complete system will be so complex that the human mind will be unable to make, in reasonable time, all of the deductions which are necessary to answer in detail many important questions. Thus, it is a reasonable conjecture that a man-machine partnership will be more effective on at least some biomedical problems than man alone. Automatic question-answering has been selected as an excellent vehicle for a powerful attack on the fundamental obstacles to a man-machine intellectual partnership.

Proposed Course of Project:

1. The development of new methods for question-answering will continue. Whenever feasible, programs will be written to test these ideas.

2. Since the PDP-10 computer is now available, it is time to integrate many of the new techniques which have been developed into a large scale, general-purpose question-answering and problem solving system. It is expected that this attempt to construct a large and practical system will bring many fundamental problems into sharp focus.

3. A biomedical data base for the large scale question-answering system will be developed in collaboration with another division or laboratory, or institute.

4. The system will be designed to be multi-purpose, so that other data bases can be added later.

5. As development proceeds, persons outside the laboratory will be invited to use and evaluate the system. These evaluations by potential users will provide valuable feedback for further improvement.

Honors and Awards: None

Publications:

Chang, Chin-Liang: Fuzzy Topological Spaces. J. of Mathematical Analysis and Applications. 24: 1968.

Hodes, Louis and Specker, E.: Lengths of Formulas and Elimination of Quantifiers I. In Schütte, K. (Ed.): Contributions to Mathematical Logic. North Holland Publ. Company, Amsterdam, 1968.

Slagle, James R.: Heuristic Search Programs. In Mesarovic, Mihajlo D., and Banerji, Ranan B. (Eds.): Formal Systems and Non-Numerical Problem Solving by Computer. In press.

Slagle, James R., and Dixon, John: Experiments with Some Programs That Search Game Trees. J. of the ACM. In press.

Slagle, James R., Chang, Chin-Liang, and Lee, Richard C.T.: Completeness Theorems for Semantic Resolution in Consequence-Finding. Proc. of IJCAI, 281-285, 1969.

Waldinger, R., and Lee, Richard C.T.: PROW: A Step Toward Automatic Program Writing. Proc. of IJCAI, 241-252, 1969.

Serial No. DCRT 6.2
1. Heuristics Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Biological Image Processing

Previous Serial Number: DCRT - 6.2

Principal Investigators: Richard Feldmann
Sam Bryan

Other Investigators: Allen Chauvenet, DCRT
Lewis E. Lipkin, M.D., NINDS

Cooperating Units: Section on Pathology, PRB, NINDS

Man Years
Total:
Professional:
Other:

Project Description:

Objectives: This project transferred to Computer Center Branch, DCRT this reporting period.

Serial No. DCRT - 6.3
1. Heuristics Laboratory
2. Not Applicable
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Automatic Pattern Recognition

Previous Serial Number: None

Principal Investigator: James R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D., Richard C. T. Lee, Ph.D., Louis Hodes, Ph.D.

Cooperating Units: None

Man Years
- Total: 1.3
- Professional: 1.2
- Other: .1

Project Description:

Objectives:

1. To develop new techniques for automatic recognition of patterns by computer in any type of data.

2. To solve specific pattern recognition problems which arise in the automatic question-answering project.

3. To solve specific pattern recognition problems which arise in the image processing project.

Methods Employed:

Theoretical analysis is the main technique used at the present time. The more promising theoretical concepts are tested by programs written for the PDP-10. The display and tablet for inputs attached to the PDP-10 allow visual on-line interaction in pattern recognition problems.

Major Findings:

Six programs have been tested and two papers accepted for publication. A program to perform automatic diagnosis on patients who have upper abdominal pain was tested on real data obtained from a Philadelphia

hospital. Diagnosis was correct in over 80% of the cases. A different program was correct in 90% of the cases. This program used a type of learning.

Image processing work includes a program to count segments of nuclei in white blood cells, a program to take sections of three dimentional objects, and a biological image software system. Some interesting theoretical results have also been obtained on the inherent logical complexity of images.

Significance to Biomedical Research and the Program of DCRT:

Pattern recognition is directly relevant to automatic diagnosis. It may also be helpful to the automatic question-answering project. It is of central importance in the processing of biological images.

Proposed Course of Project:

Work will continue using the same methods. Several theoretical concepts developed in the past year will be tested by actual computer programs.

Honors and Awards: None

Publications:

Slagle, James R.: Automatically Finding Linear Functions That Make Evaluations and Recognize Patterns. In press.

Hodes, Louis: The Logical Complexity of Geometric Properties in the Plane. Conference Record of ACM Synposium on Theory of Computing. pp. 249 - 254, May 1969.

July 1, 1968 through June 30, 1969

PHS-NIH
Division of Computer Research & Technology

Summary of Branch Activities | 1. DCRT-7

7. Data Management Branch | 3. W. C. Mohler, M.D.
Acting Branch Chief

I. SUMMARY

The Data Management Branch (DMB) came into being just prior to the end of FY 1968. The Branch represents the same diversity of functions, if not in the same depth, which characterized the Computation and Data Processing Branch (CDPB) around which the DCRT developed and from which DMB evolved. These functions include analysis, design, programming, and support for computer applications at NIH as well as some computer and punch card machine operations. In addition, DMB strives to relate these continuing functions to research and recent developments in the computer and information sciences. At the end of a year the Branch is less than a revolutionary innovation, but fortunately not a catastrophe.

II. CURRENT PROGRAMS

A. Objectives

The Data Management Branch has a primary responsibility as a central NIH resource for the application of automatic methods to the processing of the numeric and alphabetic information which are the currency of laboratory research and clinical medicine as well as of the administration of biomedical research and education. The major objective of the Branch, like that of its antecedent the Computation and Data Processing Branch (CDPB) has been to optimize the use of its limited resources for the benefit of NIH as a whole. These resources included 120 people, most of the Electrical Accounting Machines (EAM) in DCRT and a Honeywell H-200 computer which remained after the larger second generation H-800 computers were displaced in FY 1968 by third generation IBM S/360 equipment.

The major objective is divided into several specific, related objectives and responsibilities during FY 1969:

1. The Branch contains the DCRT keypunching equipment and operators. This brought with it a responsibility for keypunching all of the NIH financial management data as well as much of the computer program and data card decks from all of NIH. The objective here has been to maximize keypunch output in the face of restrictions on hiring even replacement for the normal turnover in keypunch operators.

2. The Institutes using the EAM facilities operated by DCRT in the Westwood Building had been separate from the development of computing on the main NIH campus. It was and is an objective to bring the level of sophistication of data processing in the extramural area up to that elsewhere in NIH.

3. The other EAM facilities of DCRT (Building 12) have been used less and less as part of the historic trend which began in 1958, when the first computer was acquired by NIH. The Branch carries the DCRT responsibility for providing the residual basic EAM service, an area in which it becomes progressively harder to recruit and retain competent operators.

4. Along with the H-200 computer, the Branch inherited the objective of proving or disproving the utility of this machine as a separate smaller machine to complement the S/360 computers of the Computer Center Branch.

5. During the conversion of second generation H-800 programs to S/360 programs, a separate unit was setup in the CDPB to run and maintain the converted programs. These were predominantly the programs used recurrently for data processing systems of the NIH Office of Administrative Management. The unit was taken into DMB along with the objective of enlarging the responsibility of the unit to make it more attractive for programmers.

6. The Branch includes the programmers and analysts who had previous responsibility in CDPB for applied programming exclusive of pure mathematical and statistical applications. The Branch has had the responsibility for providing such programming to NIH within the limits of its capability. It has also had the necessary objective of providing general purpose "packages" for information processing, in order to allow its own programmers as well as others at NIH to use the central computers with greatest efficiency. Thus, in addition to programming *per se* this group worked to advise, assist and teach the use of general purpose tools and to adapt to create new tools.

7. Finally the Branch has had the objective of increasing its capability as an NIH resource for the analysis of the many information processing problems which are presented to DCRT as potentially amenable to computer-based solutions, but which are too large, ill-defined or novel for other applications programmer/analyst groups.

B. Progress Toward Objectives

The new Branch was grouped into four sections, emphasizing certain functional distinctions and with these the importance of several separable methodologies, which the sections would have responsibility to develop and teach: 1) *Systems and Operations Analysis* for analyzing and specifying complex information processing requirements, *Design Methods* for analysis and design of application programs and systems, 3) *Processing Methods* using the EAM and H-200 equipment, and 4) *Systems Applications* centering around the important, if unglamorous, maintenance and recurring

production functions for systems already created. This organization has proved viable even if not the solution for all problems.

During its first eight months, the Systems and Operations Analysis Section (SOAS) was asked to analyze the information processing needs of the National Institute of General Medical Sciences (Project 7.1) to participate in the Office of Financial Management analysis of requirements for a computer system for centralized indirect cost payments on NIH grants, and to analyze the needs of the NCI for a personnel data subsystem. In addition to these larger projects, members of the Section have worked on redesign of the DCRT information and accounting systems and on three projects for the processing of scientific data in NHI, NIAMD and NINDS. The acting head of the Section serves on the NIH Committee for Extramural Program Analysis and Evaluation. There has been satisfactory progress in the application of existing analytic techniques. Much remains to be done in the explication and teaching of analytic methodologies for the benefit of DMB or NIH in general.

The Design Methods Section (DMB) continues as an NIH resource for computer analysis and programming to meet the needs of specific applications throughout NIH. It worked on more than 150 jobs distributed across nine Institutes, the Clinical Center, four Divisions (including DCRT itself), the Bureau of Health Manpower, two of the three NIH Offices of Program Planning and Evaluation, three of the four Offices of Administrative Management and the immediate Office of the Director, NIH. (Projects 7.4, 7.6, 7.7 and 7.8 are included as examples.) This represents a broad supporting service to the activities of NIH, and as such belies occasional reports that DCRT is not interested in providing service. In fact a case might be made that, in the long view, DMS, DMB and DCRT are providing too much programming service and not enough effort to encourage or compel the other Institutes and Divisions to develop programmers and computer systems analysts to support their own activities. The case here is far from black or white. At present we continue to encounter programmers in the other Institutes and Divisions who benefit greatly from the breadth of experience and expertise resident in the DMS. It remains to be shown that other Institutes and Divisions are willing or able to make a commitment to develop their own talent, except in specific instances which center around the initiative provided by specific programs or specific professionals within programs. DMB participates actively in the DCRT program of computer training and education, which in theory at least, should be increasing the number and competence of programmers throughout NIH. Nevertheless, there is room for improved educational programs, particularly at the level of advanced programming techniques which tend to be learned from the experience or preceptorial advice more readily available among the larger group of programmers in DCRT than it is in smaller groups in other Institutes and Divisions.

The Design Methods Section continues to maintain and teach the QUERY and TABLEMAKER (Project 7.2), two general purpose programming tools built specifically for NIH. During the year DMS has found it useful to "revive" and teach the use of the IBM Report Program Generator (RPG) as a tool to

aid several Institutes to make rapid progress in the use of computers for specific limited goals. All three of these tools have their limits, and DMB continues to explore other sources of tools. One source is tools existing outside NIH. There are a number of general purpose "data management" packages advertised. It is no easy matter to examine these thoroughly, much less evaluate them in light of the diverse needs of NIH and the operating rules of the central NIH computer complex. But that effort will continue.

As a result of a review early in 1968 of existing packages and tools, all of which had some undesirable features, DMB decided to use a second source, its own talent, to build its own file creation and maintenance package. (See the Project 7.3). This was a calculated risk but if successful early in FY 1970, it will give NIH a powerful general purpose programming tool to complement those already available. Another in-house tool design effort is represented by Project 7.5 (and the related 7.4). The element file-based retrieval system offers considerable promise as the prototype for the kind of system which will serve diverse NIH needs for "information retrieval."

The Systems Application Section (SAS) continues to be a success in its primary responsibility: the maintenance and operation of systems of programs which are finished, documented and in need of support for recurring production runs. The secondary managerial objective, enlarging the opportunity for members of the section to take on more advanced programming tasks, remains to be worked out.

The Processing Methods Section (PMS) has solved the problems of providing adequate keypunching for NIH in the only way possible under lowering personnel ceilings, i.e., by contracting jobs out to other parts of the Federal Government or to private firms. Some increased productivity came in-house through the installation of the "keytape" units to replace keypunch units. A broader and longer range solution of the NIH keypunching will involve the installation of modern, interactive, programmable source-data input equipment within the NIH Office of Financial Management to eliminate much of the manual coding done now by OFM. This approach is appearing widely outside NIH, but is not yet so universal as to mark NIH as completely antediluvian. The approach will require the careful planning by OFM, DMB and the Computer Center Branch.

The DMB objective of increasing the level of sophistication of data processing in extramural areas within the Westwood Building have gone well. The schedule for this effort was substantially completed by the end of 1969.

An evaluation of the utility of the H-200 computer showed it not to have clear long-term value to NIH. Its rental will be discontinued early in FY 1970 when preparations to replace all of its functions will be complete.

C. Plans

The broad objectives and responsibilities for the Data Management Branch will remain substantially the same in FY 1970 as those listed above for FY 1969. Only item 4, the utility of the H-200 computer has ceased to be an object of concern; the H-200 is leaving. Item 2, the conversion in data processing by DMB customers in the Westwood Building, is scheduled to be complete by the middle of FY 1970.

The continuing long-term objective remains: getting the greatest long-term return from DMB resources for the benefit of NIH. This is of course part of a similar, larger objective of DCRT. The short-term goals in this direction include those which will remain from FY 1969, restated in more general terms as:

1. Installation of better ways to capture and convert data to machine processable form throughout NIH.

2. Some further elimination of the dichotomy between EAM and computer technologies within DCRT and NIH.

3. Provision of more readily useful, packaged programming tools to aid the NIH staff in using the central computers; provision of advice and assistance in the use of these.

4. Increased assistance for the NIH staff in the formulation and analysis of problems for which computer processing is a possible solution.

5. Continued operational and maintenance support for appropriate sets of application programs, where this support does not inhibit the users initiative and responsibility to deal with these programs and other computer applications.

These goals are entwined with similar goals in the Computer Center Branch and the Computer Systems Laboratory, as well as with the efforts of the DCRT to provide training and education for the rest of NIH.

As part of this effort the Data Management Branch must pursue the research and developmental efforts in those parts of computer and information science which relate to the applications falling under the Branch's purview. It must establish even more effective communication with such scientific counterparts outside of NIH. A careful view of these parts of information science beyond the borders of NIH provides perspective not found within NIH or other parts of DHEW. It provides an insight into valuable areas in which DCRT should prepare itself for the future benefit of NIH. It shows the mistakes to be avoided.

Serial No. DCRT 7.1
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: NIGMS Data Management Analysis

Previous Serial Number: None

Principal Investigator: Alan K. Kreger

Other Investigators: Robert Drummond, John Rollins

Cooperating Units: The National Institute of General Medical Sciences

Man Years

Total:	1.0
Professional:	1.0
Other:	0.0

Project Description:

Objectives:

To analyze the information processing requirements of NIGMS and recommend improved methods for the management of NIGMS data. The study was undertaken at the request of and in collaboration with the Office of the Director, NIGMS.

Methods Employed:

The current NIGMS procedures for handling data were analyzed by means of data element/document/file matrices and by an extension series of interviews with selected members of the staff within the four branches of NIGMS and the Office of the Director, NIGMS.

Major Findings:

A striking finding was the degree to which each operational segment of NIGMS depended upon its own manual files for information in spite of the apparent availability of the same information from the Program Analysis Branch, NIGMS, or the Statistics and Analysis Branch, DRG. The major reason for this recurring redundancy included lack of knowledge within NIGMS of what the P.A.B. or S.A.B. could provide or previous difficulties in obtaining timely correct reports from these sources, or a combination of both factors.

At present NIGMS can, (and as a result of this study) has begun to satisfy most its operational data processing by the application of standard general purpose computer programs such as Query or Report Program Generator. In the future NIGMS will undoubtedly consolidate data and coordinate its information processing requirements, and have need for more complex analytical computer programs.

Significance:

The analysis provides framework within which NIGMS can move forward to improved organization and procedures for the use of Automatic Data Processing techniques and facilities. This was the first DCRT study of this kind covering an entire NIH Institute or Division. The findings should prove helpful to other parts of the NIH.

Proposed Course:

This project, the diagnostic analysis, will end during F.Y. 1969. Another group within the Data Management Branch will continue to work with P.A.B., NIGMS, to develop its current operational ADP capabilities. If NIGMS plans to make any major revisions of its information processing systems, it will require a longer term institute-wide effort in which DCRT may be able to participate.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.2
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: TABLEMAKER 2

Previous Serial Number: None

Principal Investigators: Gary Stoner and William Jones

Other Investigators: DMB-Unit 4

Cooperating Units: None

Man Years

Total: 2
Professional: 1
Other: 1

Project Description:

Objectives:

TABLEMAKER 2 (TM2) is a compiler level language designed and written at NIH. Its function is to simplify the writing of computer programs which create user-designed cross-frequency tables from large data files.

Methods Employed:

TM2 is a system of more than 90 Assembly Language programs. The necessity for user supplied Job Control has been eliminated by making the OS/360 Job Control largely self-contained.

Major Findings:

The current version of TM2 has been operational for almost two years. The past year again demonstrated its usefulness to NIH. The system has been modified to run under OS/360 MVT and to handle indexed sequential disk input files. These modifications have greatly increased its usefulness in the NIH computer environment.

Significance:

As a special purpose higher-level language TM2 makes it possible for programmers, and even NIH staff with no programming experience, to use a computer for studies of the frequency of occurrence of variables in large files of observations. Furthermore, instruction in the use of TM2 is valuable in helping scientists design such files. The system has been sent on request to more than twenty-five installations outside of NIH (academic institutions, other governmental organizations, etc.).

Proposed Course:

During the next year, TM2 will be extended further to increase its capabilities. The most significant change will be in allowing users to update previously created tables, which have been maintained in computer processible data sets.

Honors and Awards: None

Publications: TABLEMAKER 2 System Users Manual, Computation and Data Processing Branch, DCRT, NIH

Serial No. DCRT 7.3
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: QUADIS File Management System

Previous Serial Number: None

Principal Investigators: Gary D. Stoner, Jerry Fensterstock

Other Investigators: Robert Mount, Michael deMarco and other DMB programmers

Cooperating Units Computer Center Branch, DCRT

Man Years

Total:	3.8
Professional:	3.2
Other:	.6

Project Description:

Objectives:

1) To create a programming language, to be compiled by a system of many integrated computer programs, which will allow the scientific and managerial staffs of NIH to more easily create and update computer files 2) to extend, if feasible and economical, the language and system for similarly easy data retrieval and reporting and for creation of analytical reports.

Methods Employed:

The project represents an infrequent, if not a unique, use of facilities in PL/1 programming language for development of such a system.

Major Findings:

The project is in the later steps of system testing and debugging. Comparative studies of the efficiency and utility of the system are not yet available.

Significance:

There are many impediments to the easy use of computers by those at NIH who are not computer programmers. One of these is the process of creating, editing, and updating computer files.

The QUADIS programming language and system is designed to make this chore simpler and easier to master. It should be another step toward making the computer more directly useful to scientists and science administrators at NIH.

Proposed Course:

After the initial system has been completed, there will be a period of evaluation of the system. A decision will then be made about the appropriate direction for future extensions of it.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.4
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Computerized Bibliographic Publication System

Previous Serial Number: None

Principal Investigator: Hal Fredrickson

Other Investigators: Vivian Fykes, Louis Grosman, Myrtle Morris, Lawrence Weiner

Cooperating Units: NICHD, Scientific Information Centers Branch

Man Years

Total: 2.0
Professional: 1.80
Other: .20

Project Description:

Objectives:

To establish a generalized procedure of electronic composition of document text for medical abstract journal applications.

Methods Employed:

A generalized text-element document file format has been established. This will allow variable length document items such as authors, titles, abstracts, vocabulary, etc., to be stored as elements. A maximum of 250 elements are allowed in each document. Each element is identified by an element code. All system programs use these element codes in manipulating the element text.

Presently, all production data entering the system are in batch card mode. The batch system checks for valid element fields and the NICHD formal vocubulary items which are stored on a disk unit, and produces diagnostic messages on the computer printer. We are now developing a data entry system via terminals suitable for use by clerical personnel for remote data entry into the NICHD master file. The system will interact with the user by checking the validity of the input text and give immediate error messages for such items as missing required text fields or illegal vocabulary codes thus allowing the user to correct his errors and continue.

This system should be flexible enough so that such changes for specific users should be relatively easy to implement. Current development only supports the use of either an IBM 2741 or other terminals of similar configuration.

Once the file has been established the publication program of the system produces a magnetic tape compatable with the Government Printing Office LINOTRON electronic composition equipment. A user supplied parameter list will determine which document elements will be published and specify the page layout of the publication.

Major Findings:

1. A generalized text file format can be established that will simplify document text processing of abstract journal type applications at NIH.

2. A generalized system for electronic composition of document text from the DCRT text file format can be established.

3. A computer/user interactive input system using a terminal device such as the IBM 2741 has proven to be a reasonable means of document text input and edit.

Significance to Biomedical Research:

The publication system will enable more orderly and efficient production of "current awareness" journals thus improving communication in the medical community. The same system can with minor modifications be used for any other publications based on variable length elements within the file.

Proposed Course:

Continued development of a generalized publication system that will require minimal user effort in text input and publication specifications, taking advantage of any new hardware developments such as low cost CRT displays.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.5
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Element File Based Retrieval System

Previous Serial Number: None

Principal Investigator: Hal Fredrickson

Other Investigators: Darlene Myer, Lawrence Weiner

Cooperating Units: None

Man Years

Total:	1.0
Professional:	.80
Other:	.20

Project Description:

Objectives:

To provide NIH with an initial capability for computer-based information retrieval, making use both of manually indexed descriptor terms and of "free text" search techniques.

Methods Employed:

A user-definable, table-driven symbolic request language is used to enter retrieval requests from a remote terminal. The result of the requests are then retrieved to the terminal or the high-speed computer printer. Before each retrieval request the user has the option of defining his own symbolic names for operators, operands, variables (Element names) and macro instructions.

The retrieval system will search a text file, which has been entered in the generalized text-element format of the system, via direct access or serial access depending upon the request. A search can be performed on any text field(s) in one of 3 methods: (1) a character by character field match (2) free text search for word or word groups (3) free text "root word" search.

Major Findings:

1. A "flexible" user oriented language for text retrieval has been developed for NIH.

2. Current IBM OS/360 support for variable length direct access files is not satisfactory at present.

Significance to Biomedical Research:

The retrieval system will provide scientists and others at NIH with the facility to maintain and retrieve information describing completed work or work in progress on projects of many varieties.

Proposed Course:

Continue development of computer terminal support for the system and further develop the online data file capabilities.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.6
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Longitudinal Studies Program of Human Physiology, Biochemistry, and Psychology (Aging Study): File Development Phase

Principal Investigators: Catherine Staneck, DCRT
Arthur H. Norris, NICHD

Other Investigator: Mary Lee Dante, DCRT

Cooperating Units: Gerontology Research Center, NICHD

Man Years

Total:	.75
Professional:	.30
Other:	.45

Project Description:

Objectives:

This project is designed to be a long-term examination of the process of aging in a population of approximately 700 to 1000 men. The particular phase of the project covered within the scope of this report involves the development and maintenance of a central masterfile containing the various types of data in the study and the development of certain basic report facilities. Currently, there are twenty-six tests including physical, psychological and sociological measurements either in existence or planned for this study. Testing is done at 18-month intervals until the patient reaches the age of 70 at which time the intervals are reduced to 12 months.

Methods Employed:

With an emphasis on general applications, a system of editing and file maintenance routines have been developed to work in conjunction with those already available in DCRT program libraries. Report procedures including mean and standard deviation and regression analyses have also been prepared.

Major Findings:

This report covers just the computer related aspects of this project. The substantive scientific findings will be covered by the NICHD Annual Report.

Significance:

Prior to the inception of this phase of the project, each type of data in the study was located in physically separate and logically unrelated files with manual query and updating procedures. The creation of a single masterfile has allowed for a checking across files for consistency, for the identification and removal of incorrect data, for automatic updating of the file, and for more efficient and flexible retrieval from the file.

Proposed Course:

As the data for the remaining proposed tests become available, they will be added to the masterfile which will also be expanded by semiannual updates. Retrieval and statistical analyses across files is also anticipated.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.7
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: NIH Blood Donor File System

Previous Serial Number: None

Principal Investigator: Richard Baxter

Other Investigators: Joseph Campbell

Cooperating Units: Blood Bank Department, Clinical Center

Man Years

Total:	1.5
Professional:	1.0
Others:	0.5

Project Description:

Objectives:

The NIH Blood Bank has used a mixture of manual information retrieval techniques. These have become unwieldy, and DCRT was asked to propose and develop computer-based methods to handle the files of blood donors and of blood inventories.

Methods Employed:

The donor file programs are being written in Assembly Language to provide maximal processing efficiency.

Major Findings:

The project is still in the development phase.

Significance:

Computer processible files for donors and for blood inventory will provide, among other things, the basis for: 1) more efficient selection and scheduling of donors, 2) development of algorithms for pin-pointing donors responsible for hepatitis in the many NIH surgical cases receiving multiple transfusions (at the moment many

donors become inelligible because of accidentally having been among the donors for two patients who developed hepatitis), 3) easier and more effective control over blood inventories and records of use of blood, 4) more effective communications with donors.

Proposed Course:

The project is a collaborative effort between the Blood Bank and the Data Management Branch. When finished the system will be operated by the Blood Bank.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.8
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1968 through June 30, 1969

Project Title: Clinical Center Admissions

Previous Serial Number: None

Principal Investigator: Robert Mount

Other Investigators: Jan Polissar, M.D. and Helen DeFrancesco

Cooperating Units: Clinical Center (Jan Polissar, M.D.)

Man Years

Total: .25
Professional: .25
Others: 0

Project Description:

The purpose of this project was to design a computer-based system to maintain tape files for current and historical NIH Clinical Center in-patient records and to create daily, weekly, and monthly listings and statistical reports pertaining to Clinical Center in-patient operations.

The project is currently in the final stages of testing and will be operational before June 30, 1969.

Honors and Awards: None

Publications: None

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY
Report of Program Activities
July 1, 1969, through June 30, 1970

ANNUAL REPORT

TABLE OF CONTENTS

Page

DCRT Annual Report
July 1, 1969 through June 30, 1970
Office of the Director

The Division of Computer Research and Technology is charged with insuring that computers become a useful, creative part of the research, education and management programs of the National Institutes of Health. The variety of activities among these programs required the development at NIH of what can properly be called a "computing manifold": an entity consisting of and at the same time uniting, many diverse elements. We believe that the term is preferable to the much used and abused word "system", which has suffered a semantic inflation. It also implies a set of related elements but usually connotes an attempt at uniformity which often fails because of the very variety with which it must cope. There are separable but interconnected parts within the DCRT manifold, which in total consists of research, development and service as these in turn relate to computation and information processing throughout NIH.

These parts are most easily understood when the separation is made first in terms of the computer (hardware/software) systems developed and interconnected by DCRT, and second in terms of the programs and projects of the Laboratories and Branches of the Division.

But before entering into the details of this analysis it is important to emphasize several salient facts:

1. The DCRT is, for practical purposes, only four years old. Though (legislatively) established in FY'64, the Division was not fully activated until August '66 (FY'67). Recruitment of necessary staff and design and implementation of a new computer configuration began at that time.

2. The DCRT has fortunately been able to recruit and retain an excellent cadre of scientific and technical expertise, thus permitting both the development of the sustained interdisciplinary program which was expected of it plus the achievement of many solid individual accomplishments.

3. Progress over the past three years and particularly in FY 1970 have placed NIH in an unparalled position of excellence in computing. The NIH Computer Center, developed and operated by the DCRT Computer Center Branch, is without peer in DHEW and probably is one of the better biomedical computer facilities in the world. Computer systems implemented in the NIH laboratories, have been in the forefront of service to the biomedical community. The DCRT provides a very competent (though certainly not the largest) group of applications programming talent within the DHEW.

4. Overall, computing at NIH has made rather significant strides in becoming a useful, creative part of the programs of the NIH,

even though there is much room for improvement of what exists and for new advances in many areas of management and biomedical science.

5. Almost two-thirds of the DCRT staff operates under the NIH Revolving Fund and provided service to every major organizational segment of NIH in FY 1970. The occasional complaint that DCRT is more interested in "research" than "service" unusually arises when it is unable because of personnel shortage to respond with what a customer feels is ideal service for his specific needs. Fortunately, even these incidents have decreased during the last year.

The Computer Systems

The computational capacity is distributed among three kinds of computer systems. The first category is the large capacity central NIH computer facility in Building 12. This is composed of three interlinked IBM-360 processors which work on a single, shared computer program and data set library; a new technology pioneered by the DCRT. This facility, described in the Computer Center Branch summary, processed more than 1800 jobs per day for 800 users in FY 1970, compared to 750 jobs per day for 600 users in FY 1969.

This striking success is attributable, in large part to two software facilities WYLBUR and CPS (Conversational Programming System) and to the rapid, effective access provided by terminals in or near the laboratories and offices of the users, connecting as needed by standard telephone lines. The success is also attributable to a conscious effort to make the NIH computing center a practical utility for the largest possible number of users if not all things to all people. The center runs under standard operating procedures fully described by the Computer Center Branch in its "User's Guide". The central system is well supported by an extremely competent set of creative systems programmers, and a dedicated operations staff.

The second category includes computers designed for and shared by specific groups of users whose immediate needs require and can be met most effectively and efficiently by machines of smaller capacity located in close proximity to their place of work. To date most of the systems have been developed for laboratory research in the Institutes, namely NIAID, NIDR, NICHD, NHLI and the part of NIMH in Building 10. They are, in general, noteworthy for their ability to capture, format, store, interpret and display experimental data from separate sources for several users in the same time frame. They are discussed in the summary of the Computer Systems Laboratory, which provided the expertise for their design and implementation.

The DCRT itself has computers in this second category. The PDP-10 computer within the Computer Center Branch (CCB) is able to serve a striking variety of users within DCRT and in other parts of NIH by virtue of its time-shared operating system, the other software facilities which have been implemented, and its graphic input (RAND tablet) and output (cathode ray tube) devices. The Computer Systems Laboratory (CSL) has developed a SEL-810B computer into

a prototype system which makes the touch-tone telephone an effective means of communicating with other computer. The SEL-810B contacts and connects with those computers and translates the digital response from them into voice messages back to the user at his telephone.

The third category is composed of small capacity, dedicated machines each of which has been developed to serve a specific purpose in contradistinction to the two previous groups in which each machine is used for a variety of applications. Examples include a system being designed for preprocessing EKG signals and another to handle NMR Spectra in Building 2.

Together these systems provide the basic capability for NIH to acquire and transform information into machine processable form, to manage this data within the computers and to retrieve and analyze it, thus providing the transformations or translocations by which the information becomes useful to scientists and managers.

Once having designed the three categories of computers, their interconnection became the second order of business. The main purpose for such connection is access to greater capacity as needed and when appropriate without having to provide this capacity as an overhead cost for the smaller machines. The design philosophy for such connections has anticipated the use, wherever possible, of standard telephone lines, connecting only as needed. A second purpose for interconnection is the capability to access or transfer information (programs as well as data) and thus to make the accumulated knowledge which they represent rapidly useful in more than one computer. The overall purpose is the creation of a flexible network which can be created and connected by the users as they need specific facilities and capacity. Suffice it for this report to note that the PDP-10 already links to smaller machines and that the Computer Center Branch has completed an operational version of the Data Store Program which will provide access for machines in the second category to and from the larger capacity of the NIH computing center.

It must be noted that the time frame required for the design, development, implementation and effective utilization of computer hardware/software systems usually extends well beyond a single year: Mean development time for such facilities is three to five years. In retrospect we can report that design development and implementation was largely completed within three years, and that this past fiscal year has been a test of the design philosophy and operational capability. The test has proven successful since the Computer Center, operating on a fee-for-service basis, has fully recovered its operational personnel and supply costs by charging the very low resource-hour rate of $175.00. A modest reduction to $162.00 per resource hour will be in effect during FY'71.

Computer systems are never static. Their further evolution, however, depends on the changing needs of the substantive program. The NIH now has a remarkably versatile computer capability. Further development will require program scientists and managers to acquire additional expertise in computer utilization and to express fully their program needs.

The DCRT Programs

The Division's programs have three primary activities and outputs: provision of computer facilities, development of systems, and research. They yield secondary outputs which are also important for other parts of NIH: advice and information, and trained staff. The summary for each Laboratory and Branch describes its specific focus and the ways in which it makes use of the Division's basic disciplines: computer science, engineering and mathematics.

The DCRT research program involves staff from all its Laboratories and Branches, individually, in concert with members of other parts of the Division, or in collaboration with scientists from other NIH Institutes and Divisions and from other research centers outside NIH. One segment of the DCRT research is "intensive", i.e., its objectives lie within the basic DCRT disciplines; this can be considered the "fundamental" DCRT research. The work deals with abstract concepts in mathematics, logic, and linguistics, or with basic methods (algorithms) which derive from these for use in attacking general classes of problems, or with studies applying those methods to bodies of experimental data. The Heuristics Laboratory (HL) devotes much of its time in this area. Other work goes on in the Computer Center Branch, the Computer Systems Laboratory, the Laboratory of Applied Studies, and the Physical Sciences Laboratory, but in the latter two laboratories most of the development of theory and method arises as part of a second kind of research.

This second variety of research is "extensive". It is an outward extension of the basic DCRT disciplines, and its goals are the discovery of new biomedical knowledge. Within DCRT, physicians, biologists and physical scientists, whose competence in their special areas is supported by a knowledge of computation, work effectively and productively with statisticians, programmers, and mathematicians, whose disciplines supply the formalisms necessary for the processing and analysis of data. Most of it involves collaboration with groups outside of DCRT. The work covers a wide range of biomedical subjects: analytic biochemistry (including virtually all forms of spectroscopic analysis), the physical chemistry of liquid states, macromolecular conformational chemistry, kinetic theory and enzymology, membrane transport, molecular mechanics, molecular biology, neurophysiology, cardiac physiology, immunology, pharmacology, cardiology, clinical pathology and psychometrics, and taxonomy. This kind of research appears, as expected, among the projects listed for the Laboratory of Applied Studies and the Physical Sciences Laboratory. Similar projects also appear under the Computer Systems Laboratory, the Data Management Branch and even in the Computer Center Branch, emphasizing the fact that the proper primary focus in DCRT for "extensive" projects cannot be anticipated in advance.

The systems development activities of the DCRT may appear confusing to the outsider. This impression is correct and understandable: the activities reflect the diversity of programs at NIH which, in turn, dictate the continuum in DCRT. At one extreme the simplest system is a single computer program with its attendant control language developed for a single special application of a single user on an existing, operational computer. At another extreme is the development for multiple users and uses of an entire computer system composed of hardware, systems software and multiple special

application programs. The development of them has already been mentioned above under "The Computer Systems." In a third direction lies the development of general purpose tools and facilities for an existing computer system to similar applications of many users or to facilitate the development of specific applications programs of such users. The DCRT staff contributes expertise in statistics, mathematics, and problem analysis, thus making it possible for NIH scientists to extract information from their data by utilizing quantitative methods. The DCRT applications programming expertise also aids the NIH administrative and managerial functions. The work in this third DCRT activity is done on a fee-for-service basis and is cited in the DMB and LAS reports.

Only by devoting part of its resources to developing or to acquiring and adapting and then to maintaining, using and teaching the use of new tools and facilities, can the DCRT make it possible for the limited DCRT staff to provide service for more applications, and make the computer useful, and accessible to the NIH staff. This philosophy has proven effective in that there has been a great increase of service to users without increase in DCRT staff. Any alternative would require the NIH staff to become expert applications programmers or to be dependent forever on programmers in DCRT or elsewhere to get their work done for them.

The DMB, LAS, CSL, and PSL also provide consultation, advice, and assistance for those throughout NIH who want to develop their computer solutions but need help in choosing the best approach. This is then a fifth kind of activity for the DCRT staff. The Computer Center Branch also advises, and consults on the use of the central computer system and on other subjects which are the specialty of members of the branch.

A sixth DCRT activity consists of providing information and education about existing DCRT capabilities and facilities, about the means and ways in which to use them, about the whole spectrum of activities within the Division, and about the basic disciplines behind computing. The program of training courses and seminars developed and coordinated by CCB has been, in fact, taught by a faculty from all Laboratories and Branches, even though this is not obvious from their reports. The courses and seminars are well subscribed and provide multiple benefits for DCRT and NIH: the courses are well taught, are relevant to the NIH computing systems, are a stimulus to the DCRT staff to produce more responsive facilities, and insure that the "students" become familiar with DCRT and its staff when they become familiar with the computing systems. The training courses are backed by the Technical Information Office, CCB.

Other educational and informational activities are located within the Office of the Director, DCRT. The DCRT Library is a unique resource to NIH and others, as well as an essential resource for DCRT. The DCRT Scientific and Technical Information Officer joined the staff in the latter part of fiscal year 1970 with an unusually appropriate background in biomedicine and information science. The entire education/information program has developed very well over the last year.

The Prospect

The choice of the term "manifold" for computing systems at NIH and for the DCRT itself, risked raising familiar images of something containing nothing more substantial then either a vacuum or hot gas. Both the computer systems and the Division programs can tolerate the risk; there is solid substance throughout. There is, of course, still much to be done.

The diverse elements of the computing systems must be connected and used more effectively as well as efficiently. The elements of the DCRT can be still more effective, particularly as they pursue the Division's purpose to create at NIH knowledgeable computer usage, by independent mature users of the facilities which DCRT can provide.

Finally, there is a larger set of elements of which DCRT is only one. This is the entire NIH community of computer users and information processors, present and potential, individual and organizational. DCRT will continue to be a necessary (if not the singly sufficient) part of the NIH manifold, so that NIH may use computers effectively and efficiently, and so that the NIH staff incorporates mathematics, statistics, and information science appropriately (as well as aptly) into the increasingly complex substance of modern biomedical research and education and its administration.

July 1, 1969 thru June 30, 1970

PUBLIC HEALTH SERVICE-NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

1. DCRT-2
Serial Number

2. COMPUTER CENTER BRANCH

3. J. D. Naughton
Branch Chief

I. SUMMARY

The Computer Center Branch is responsible for designing, planning, implementing and maintaining a large general-purpose computer facility to meet most effectively the dynamic and diverse requirements of both N.I.H. research investigators and managers in the support of modern medicine. This charge includes the development of computer tools that will bring the computer to bear on problems at every level of research and in many locations. The core of this computer facility is a network of computers and terminals, which, by means of modern communications techniques, makes the power of the computer immediately available in all laboratories and offices throughout N.I.H. An inherent part of this network is the continued research into extending the network into new labs and areas where the computers can assist medical researchers in their mission. The introduction to, and educating of, new personnel in the understanding and use of computer tools, goes hand-in-hand with the development of any new computer techniques.

A full spectrum of computing power is to be provided to research personnel. Large computer systems as well as mini-computers and terminals will be lashed together providing capability at many levels. Research into computer and information sciences coupled with the N.I.H. medical investigators' needs and cooperation will bring computers closer to the research environment where they can perform most effectively in attacking the complex problems of modern medicine.

The medical research programs of N.I.H. require the most powerful and flexible of computer services and tools available today. The computer network provided must have a distributive power that is easily accessible on demand to scientists in the research environment. The computer is ours, to mold, polish and, in general, enhance into a complete tool for medical research and its administrative support. New areas of computer applications will be sought out continuously, in conjunction with a comprehensive program of educating new and old computer users in how to use computers most effectively in their domains.

The past year has been fruitful; the computer network has become a reality. Ten 360/20 computers and over one hundred selectric typewriter terminals have been installed and are in productive use by over 900 users at N.I.H. Additional types of terminals for this network are being planned and tested. The art of communicating between computers of different types and manufacture has progressed at a satisfactory rate. Protocol for intercomputer communication has been established and signal exchanges between systems inaugurated. The 360 processors have communicated with the AGT-30 and P.C. (Programmed Console) systems, while the PDP-10 has conversed with a PDP-12, LINC-8, and SEL 810 mini-computers. Thus the power and facilities of the large central computer systems are available to the users of small laboratory computers at their own locations. Mass storage, high-speed line printers and access to more powerful computing capability are some of the features available to the small systems. In essence the full power of the large systems is available to the small computers for the asking by the addition of a telecommunications capability. This means that the investigator with the lab computer has all that power available to him without the inherent additional costs of such facilities.

Two interactive terminal systems (CPS and WYLBUR) offered by the Computer Center have met with overwhelming acceptance. CPS, the Conversational Programming System, is a multiple user, interactive computing system, while WYLBUR is an on-line text-editing and remote job entry system. Nearly 200 users have registered their selectric typewriter or teletype terminals in order to use the system. The demand for service is increasing at such a rate the Center is working with the phone company to make 25 lines available to the CPS system, and 80 lines available to the WYLBUR system. These terminal systems are used in many different facets of the work done at the N.I.H. One notable use of these systems is the Blood Donor System set up in conjunction with the Clincal Center Blood Bank. Both terminal systems are used to maintain and use the blood donor file. Compatable blood donors can now be found easily, but more important, quickly. This system has allowed N.I.H. to eliminate its use of commercial blood completely, thereby reducing the probability of serum hepatitis by 90%. The system makes possible more detailed typing and selection of donors.

The use of computer driven graphic displays as a research tool continues to grow. The steadily increasing utilization of the graphics devices attest to their acceptance by the research community. The success of the graphic systems on the PDP-10 and 360 computers is a tribute to the monitors and software packages developed to enable a researcher to use the displays without being a hardware or software expert. This man-machine interaction enables the user to

manipulate graphic data under his control and visually watch the effects of the parameter variation on the biological or other system. The graphics work done was honored in December when the Computer Center Branch hosted a meeting of SIGGRAPH, the special interest group on graphics of the Association for Computing Machinery. Over one hundred persons attended the meeting which had the theme of "Graphics for Modern Medicine."

The molding of our computer operating systems to remain sensitive and responsive to the research community's needs continues. The latest pertinent computer science techniques are employed to maintain or reduce response time while servicing an ever-increasing number of users. Both the 360 and PDP-10 systems have been upgraded to maintain the quality of service while adding new features. A comparator has been designed and built for the PDP-10. The comparator plus associated software provides light-pen capabilities to the rand tablet. The 360 Remote Job Entry network has been expanded to 10 systems, while 3 or 4 more are in the planning stages.

Computer Center users are kept informed of all activities through INTERFACE. Subscriptions to INTERFACE now number over 1500 copies per issue with many copies being distributed throughout the country. The Computer Center Users Guide has been much in demand as a model of communication between computer centers and their users. The expertise in the user support area has increased. The success of work in this area is attested to by the wide acceptance of this group by users of the Computer Center.

The utility to handle paper tape was completed. Lab scientists can now record their data on an inexpensive medium and still be able to have this data processed by the central computing facility. Of more significance is the newly established ability to process analog signals on-line. For the first time up to 16 channels of analog data can be accepted, by the computing system, and converted to digital form. The data can then be processed by library or user supplied programs. This facility will allow N.I.H. scientists to apply the power and flexibility of the digital system to the analysis of continuous biological functions such as blood flow, blood pressure, heart beat patterns and brain waves. Data from ultra centrifuges and radioisotope scanners can now be interpreted with the many intermediate and time consuming steps presently required. This facility currently analyzes 20-30 EKG's (electrocardiograms) per day of patients in the N.I.H. Clinical Center.

Now that these two services are operational other methods of collecting and inputing data are being investigated. A service involving an optical page reader is being designed and is scheduled for installation next year. Such a system should enable investigators to design forms for mass data collections that can then be processed by the computer directly instead of undergoing one or more steps, like keypunching, that they currently go through. Large surveys will be more easily handled and the time lag between the end of collection to interpretation of data reduced significantly.

The processing of on-line files is in the final stages of implementation. This on-line system will allow a user to maintain, retrieve and display information in a data base via a remote terminal. Thus a powerful means of accessing critical information quickly will be at the investigator's command. When completed this will make three different remote systems that the user can employ to retrieve or massage on-line information.

The direct application of computer techniques to biological research in collaboration with N.I.H. scientists led to the development of methods of man-machine interaction which permits the trained biologist to represent to the computer the relationships between components of biological images. Using syntax-directed picture processing techniques, algorithms were developed to efficiently process digitized images. Similarly, data collected by a high-resolution mass spectrometer has been analyzed to determine possible paths of molecule decomposition. Reconstruction of molecules is done by combinatorial methods, showing that FORTRAN can be successfully used for topologically oriented problems.

The program to educate scientists and administrators in the use of information science techniques and the Computer Center continues to grow. Over 35 courses are offered two or three times a year. Over 1200 registrations were handled in the past year while several hundred others could not be handled due to lack of sufficient staff and space. This program has made a significant contribution to the application of computers to biological research by making the users more aware of how to use the the computers and how to apply information science to their areas.

Serial No. 2.1
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: User Support and Communications

Previous Serial Number: Same

Principal Investigator: Frances E. Halverson

Other Investigators: Staff of PAL Unit

Cooperating Units: None

Man Years

Total:	7
Professional:	7
Others:	0

Project Description:

Objectives:

To provide the users of the Computer Center with the personal assistance necessary if they are to make effective use of the Center's facilities. To provide users assistance in resolving problems encountered while using the systems (hardware and software) maintained by the Center.

Methods Employed:

The PAL (Programmer Assistance and Liaison) Unit was established to perform this function.

Significance to Program of the Division:

The computer user at NIH has at his disposal a group of competent professional programmers that are able to assist him with his problems in running programs or

using the computer facility. The PAL Unit notes all trouble areas and, through INTERFACE, communicates common problem areas to all users. In all facets of computer use the user has someone that can give him the answer or guarantee to find the answer for him.

Honors and Awards: None

Publications: None

Serial No. 2.2
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer Center Users Guide

Previous Serial Number: Same

Principal Investigator: Frances Halverson

Other Investigators: Leslie Barden, PAL Unit and Systems Team

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

To provide the users of the central computer facility with a guide to the services, standards and use of the Computer Center.

Methods Employed:

A 200-page Users Guide was published and distributed to all users of the Computer Center, to Institute and Division Administrative or Executive Officers, to contract companies required to use the Computer Center and to other organizations and individuals having a logical need for it. Updates to the Users Guide are published and distributed as necessary to keep it current.

Significance to Program of the Division:

For the first time all information pertaining to computing was brought together in a single reference document. The Users Guide contains a description of all facilities and services and how to use them. Programming standards, languages supported, JCL summary and other facets of computing are all given in detail. It is a complete guide to computing for the computer user.

Honors and Awards: None

Publications: None

Serial No. 2.3
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Paper Tape Facility

Previous Serial Number: Same

Principal Investigator: Robert H. Brunelle

Other Investigators: S. Lyles

Cooperating Units: None

Man Years

Total: 1/4
Professional: 1/4
Others: 0

Project Description:

Objectives:

To provide a means by which investigators, whose laboratory gear produces paper tape, can have the information on these tapes processed by the central computing facility.

Methods Employed:

A paper tape reader and punch have been acquired and put on the 360 computer. A general-purpose program is being written that will transfer information from paper tape onto 9-track magnetic tape. The data can then be analyzed by 360 computer programs.

Siginficance to Program of the Division:

Laboratories will be able to acquire less-expensive gear on which to record their data, and still have the data analyzed by the central facility's computers.

Honors and Awards: None

Publications: None

Serial No. 2.4
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: HASP - Shared Spool Disks

Previous Serial Number: Same

Principal Investigator: John Gilman

Other Investigators: None

Cooperating Units: None

Man Years

Total: 1/2
Professional: 1/2
Others: 0

Project Description:

Objectives:

1. To provide a single logical job input/output stream for all three IBM 360/50's (future 360/65) at DCRT-CCB.

2. To allow for more dynamic balancing of the work load among the multiple computers in the installation.

3. To provide an overall smoother work flow within the Operations Section of the branch.

4. To facilitate information gathering about the current status of work on the three systems.

5. To optimize the use of on-line storage devices for intermediate storage of job input and output.

Methods Employed:

The HASP spooling system currently in use on the 360/50's is to be modified to allow all three copies to use the same set of on-line spool disks.

Significance to Program of the Division:

The multiple computer environment at the Computer Center allows significantly better overall service to the NIH computing community by providing both increased capability and superior reliability to a single machine operation. Within this framework, however, the task of managing three computers rather than one presents formidable scheduling and logistical problems. Through the use of a shared input/output queue, many of the decisions currently being made manually can be automated, thus providing more immediate and accurate response to the variations in work load and resource availability.

Honors and Awards: None

Publications: None

Serial No. 2.5
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: 360 Systems Development

Previous Serial Number: Same

Principal Investigator: Robert H. Brunelle

Other Investigators: Staff of Systems Team

Cooperating Units: PAL Unit

Man Years

Total: 6
Professional: 6
Others: 0

Project Description:

Objectives:

To maximize the thruput and minimize turnaround time to all users of the central facility's 360 computers. To put computing power into every researcher's lab or office thru remote terminals and software systems. In general, provide the NIH user with the best Computer Center and service he can get thru software development and hardware expansion.

Methods Employed:

Judicious selection and tailoring of software systems to the NIH environment. Acquisition and development of software and hardware to provide the researcher with the tools he needs. Constant attention to overall system software and hardware needs so that new systems and hardware are available when needed. Increasing the capacity of the system to keep ahead of the needs of the NIH users.

Significance to Program of the Division:

Constant attention to all facets of computing provides the NIH with an up-to-date computer facility catering to the needs of all NIH's researchers. Computer power in the lab brings a powerful research tool closer to the project, thus making it easier to use, and more likely to become an integral part of, the research program. All of these efforts combine to form a forward looking Computer Center sensitive to the needs of the NIH research environment.

Honors and Awards: None

Publications: None

Serial No. 2.6
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: WYLBUR Text-editing System

Previous Serial Number: Same

Principal Investigator: Roger Fajman

Other Investigators: J. Smith, L. Barden

Cooperating Units: None

Man Years

Total:	3/4
Professional:	3/4
Others:	0

Project Description:

Objectives:

1. To provide Computer Center users with a convenient and comprehensive tool to aid in the creation and development of programs.

2. To allow, secondarily, for the creation and editing of text materials such as letters, reports, etc.

3. To provide a convenient method for ascertaining the status of the computing system as a whole and locating particular jobs as they are being processed.

4. To lower the overall volume of materials which are handled by the Production Unit directly.

Methods Employed:

WYLBUR is an on-line text-editing and remote job entry system oriented to low-speed character mode terminals such as the IBM 2741, 1050, and 2260 and Models 33, 35 and 37 teletype machines. It was originally designed and implemented at the Stanford University Computation Center. WYLBUR provides the user, in his home or office, the facility to create and edit source programs in real time, submit them for compilation and execution by the standard job stream processor, and retrieve the results of execution at his terminal. In addition, the user may work with arbitrary text material, libraries, and interrogate the system about the current status of the batch job stream.

Significance to Program of the Division:

WYLBUR represents a significant move towards lowering the amount of non-programming overhead involved in the development of programs. With the job output available at the user's terminal, effective turnaround time is substantially lowered. In addition, the added ease of using a typewriter-like terminal instead of a keypunch contributes to more rapid development and debugging of programs.

Honors and Awards: None

Publications: None

Serial No. 2.7
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Lavender Box

Previous Serial Number: Same

Principal Investigator: Robert H. Brunelle

Other Investigators: E. Alterman, J. Camp, R. Fajman

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

NIH is acquiring a large variety of terminals that will need to communicate with the central facility. This project addresses that problem. It is an attempt to outline a plan for DCRT to establish a standard communication language that will enable any type of terminal to communicate with the central system. Further, it describes what the central facility should look like, and how and what type of services the Computer Center will be able to offer the users of remote terminals.

Methods Employed:

The plan for DCRT's implementation of such a project has been named the Lavender Box Project. The prime function of the Lavender Box Project is to provide total systems control for a large central computing complex surrounded in the periphery by multitudinous

terminals. The types of terminals vary from simple typewriter terminals to small computers. The functions performed by these terminals also vary widely from conversational mode terminals, RJE, to requests for computer power or data storage.

Significance to Program of the Division:

The most important aspect of the NIH terminal system is that an NIH standard communications language be defined. Once this is done, any type of remote will be able to use the Computer Center's facilities.

Such a system will give the Computer Center, DCRT, and NIH a powerful and flexible system. The capacities of the central system will be available to remote locations, all systems will be more reliable, and the computer load will be balanced dynamically rather than manually.

Honors and Awards: None

Publications: None

Serial No. 2.8
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Data Store Program

Previous Serial Number: Same

Principal Investigator: James Oberthaler

Other Investigators: William Jones

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

The primary purpose of the data store system is to supply the equivalent of a large on-line data and information storage capability to users of small computers outside the central computer facility. This capability will vastly increase the utility and power of the many small computers in the various laboratories and clinics.

Methods Employed:

The small remote computers will converse with the data store system via communication lines. They will request various activities to be performed with regard to a large data file. Capabilities include reading of data currently stored in the file, creating new files, duplicating files and selective processing of subsets of this data.

Significance to Program of the Division:

This is an important step in the Computer Center's plan to create an environment in which the many small specialized laboratory computers can easily converse with the powerful central facility. Effectively it will enable the researcher to have a large computer at his finger tips with the resultant computational and data processing power available locally to the lab, clinic or office.

Honors and Awards: None

Publications: None

Serial No. 2.9
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Pharmacology Project, AGT-30

Previous Serial Number: Same

Principal Investigator: Charles R. T. Bacon

Other Investigators: Joel Kaufmann

Cooperating Units: DRFR

Man Years

Total: 2
Professional: 2
Others: 0

Project Description:

Objectives:

The system is being developed collaboratively by DRFR and DCRT as a research tool for pharmacology and toxicology. The project is a first step in DRFR's development and evaluation of an integrated set of computer techniques for handling data on chemical compounds and the functioning of living systems. DCRT is structuring its effort so that the computer systems developed can be generalized to support other areas of biomedical research and a wider community of users at NIH.

Methods Employed:

The terminal consists of a computer with CRT, designed to display three-dimensional picture information with great speed. It displays two-dimensional projections of the three-dimensional information which is stored in its memory, and the user is free to alter viewing angles, translations, and magnifications. Telephone data communication problems, e.g., crystallographers or pharmacologists can use the terminal to observe models, and compose new calculations to feed to the 360.

In addition, to work toward the goal of accessing 360 data, and providing a viewing service for FORTRAN and PL/1 programmers, work is nearly finished on a sketching program designed to permit input of chemical structures by means of drawings. Double, triple, high-energy, and steric bonds can be specified, and a large number of atom types can be represented. The sketch, as drawn, is made to correspond to a connection table in which all relevant structural information is contained. Work proceeding on the PDP-10 and elsewhere by other groups, for translating connection tables to Wiswesser notation, relates to this in a way which suggests a rather powerful chemical structure input system.

Finally, a few workers have experimented with applications involving tape input to the AGT-30. Although not as convenient as telephone data communication, this nevertheless provides a good viewing service.

Honors and Awards: None

Publications: None

Serial No. 2.10
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Total: Analog Data Input to the Central Computers

Previous Serial Number: Same

Principal Investigator: Joseph D. Naughton

Other Investigators: James Bailey, Sandy Liles

Cooperating Units: None

Man Years

Total:	1
Professional:	1
Other:	0

Project Description:

Objectives:

The Computer Center, DCRT, has acquired an IBM 1827 Data Control Unit which is able to digitize up to 16 channels of analog data. This device permits investigators with analog data to convert it to a digital form for processing by library- or user-supplied programs. The 1827 will accept analog data physically brought to the central facility on magnetic tape.

Methods Employed:

The system is currently in a limited production environment and enables the NHLI to convert electrocardiograms on 1/4" analog magnetic tape to 9-track digital tape for further analysis and diagnosis. Tapes submitted by 5 p.m. are converted by 8 a.m. the following morning. NHLI and LAS are currently working on putting the diagnosis on CPS data sets for retrieval by 2741 terminals at the Clinical Center.

A half-inch instrumentation tape recorder will permit users with analog data on IRIG standard tapes to submit their analog data along with their jobs on a full-service basis. The data tapes will be digitized and 9-track digital tapes or disk data sets will be created. These data sets will then be available for subsequent processing by the user's program.

The 1827 will convert analog data into a fixed point half-word format with either 8-, 11-, or 14-bit resolution. The instrumentation tape recorder has FM electronics for all of the standard speeds between 1-7/8 ips and 60 ips, inclusive. In addition, direct electronics for all speeds up to 120 ips is available for one track. The first six tracks are reserved for data, while the seventh track contains a standard IRIG B time/event code which is used for controlling the digitalization of the data. This unit should be in service by the summer.

Significance:

The ability to process analog information on the NIH central computer system will allow NIH scientists to apply the power and flexibility of the digital system to the analysis of continuous biological functions such as blood flow, blood pressure, heart beat patterns and brain waves. It will also enable them to interpret data from ultra-centrifuges and radioiosotope scanners without the necessity of the many intermediate and time-consuming steps presently required. This allows a great deal of clinical and research information to be analyzed which was simply ignored previously because it was impossible to process the massive amounts of information available in any reasonable amount of time.

Honors and Awards: None

Publications: None

Serial No. 2.11
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: RJE System

Previous Serial Number: Same

Principal Investigator:

Other Investigators: Staff of Systems Team

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

A remote Job Entry System (RJE) using IBM 360/20 terminal computers will allow any 360 job to be transmitted from a remote location directly to the central 360 system to be processed along with other tasks, and the results can be transmitted back to the remote location for printing. The RJE user is provided the full power of the large central 360 system at his own location including all standard programming languages, libraries, catalogued procedures, and utilities normally available in the central system when 360 jobs are submitted at the Computer Center. Use of the RJE terminals requires no programming or job control language changes. RJE terminals installed in the Institutes and Divisions will accept both express and checkout type work. There will be, however, a limit on RJE input and output volumes.

Methods Employed:

The RJE terminal consists of three inter-connected units: (1) 600 card-per-minute reader; (2) a 300 line-per-minute printer with 132 print positions using a 64-character print element and a controller/transmission interface. The interface is connected to a "data set" (telephone company modulator/demodulator) which links the terminal through a standard telephone line voice grade, 2400 baud, directly to the central 360 system in Building 12.

Significance to Program of the Division:

The RJE service has had a significant impact on reducing the elapsed time between project design and actual implementation. The rapid service provided via the RJE terminal system has reduced the time required for testing and debugging new programs and algorithms from weeks or months to days or, in some cases, to mere hours. This has not merely allowed projects to be completed faster, but, because of the resultant manpower savings, has allowed many additional research programs to be conducted which would not have been possible otherwise.

Honors and Awards: None

Publications: None

Serial No. 2.12
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: INTERFACE (technical newsletter)

Previous Serial Number: Same

Principal Investigator: Joseph D. Naughton

Other Investigators: Staff of Systems Team and PAL Unit

Cooperating Units: None

Man Years:

Total: 2
Professional: 1
Others: 1

Project Description:

Objectives:

1. The primary purpose of this effort is to give the Computer Center a regular means of conveying, (1) highly current technical information to the widely dispersed computer programmers and systems analysts at NIH, and (2) current Center accomplishments, plans, policies, course offerings and other information of interest to personnel and general managers at NIH.

2. A secondary purpose is to foster communication and collaboration among computer users and between them and the Center by providing a forum for: (1) announcing key new projects, accomplishments, personnel or organizations; and (2) airing viewpoints or suggested approaches to computing problems.

Methods Employed:

INTERFACE is published every three-to-four weeks, and distributed to all scientific and administrative personnel who have expressed a desire to be kept up-to-date on computing at NIH.

Significance:

INTERFACE complements the _Users Guide_ and other technical manuals by highlighting items and directing readers to the other publications for details. INTERFACE has given the users of the Computer Center a single reference point for all communications concerning the use of computers at NIH. It has done an excellent job of keeping the users informed of new services, major systems changes and all facets of computing at NIH.

Honors and Awards: None

Publications: None

Serial No. 2.13
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: IBM 2250 Graphic Program Support

Previous Serial Number: Same

Principal Investigator: James Oberthaler

Other Investigators: Jennifer Smith, Robert H. Brunelle

Cooperating Units: None

Man Years

Total:	1/2
Professional:	1/2
Others:	0

Project Description:

Objectives:

The objective of this project is to design, implement and maintain a PL/1 Graphic Subroutine Package to enable NIH users to communicate with the IBM 2250 Graphical Display using OS PL/1.

Methods Employed:

The PL/1 Graphic Subroutine Package is designed to enable the programmer to communicate with the IBM 2250-1 using OS PL/1, version 4. The programmer using this package can have access to the 2250-1, as well as the resources of PL/1. The ability to issue calls to procedures written in assembly language enables the programmer to send data to and from the display. Interrupts are handled by user-defined ON conditions. The subroutines included in the IBM System/360 Operating System Graphic Programming Services for the IBM 2250 Display Unit, Form C27-6909, are also made accessible to the programmer.

Significance to Program of the Division:

NIH users will be able to extend the data processing power of System/360 computers: (1) to handle the graphic information associated with medical research and analysis applications; and, (2) to provide faster and more effective retrieval and graphic expression of medical data.

Honors and Awards: None

Publications: None

Serial No. 2.14
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: CPS - Conversational Programming System

Previous Serial Number: Same

Principal Investigator: John Camp

Other Investigators: Trudy Kenny

Cooperating Units: None

Man Years

Total:	1-1/2
Professional:	1-1/2
Others:	0

Project Description:

Objectives:

CPS continues to provide much of the power of the central computer facility to the researcher in his lab or office. By using CPS the researcher has available the computing capability of the central 360 computers as well as the ability to access data bases stored on-line at the central site.

Methods Employed:

The CPS system allows multiple users to simultaneously write, debug, and execute 360 computer programs conversationally. Programs are written in a subset of the PL/1 language from a 2741 typewriter terminal or teletype located in or near the user's office. The terminals are connected with the NIH central 360 system over telephone lines on a dial-up basis. In addition CPS supports a dialect of the BASIC language and provides a remote job entry (RJE) facility through which jobs can be submitted to be run in the background batch environment.

Significance to Program of the Division:

CPS continues to provide computing capacity in the lab or office allowing the researcher to utilize the central computer facility without leaving his normal working environment. Thus the computer is more readily available and much time is saved by not making the user come to the central site. Continuing enhancements to CPS place increasing power at the user's disposal. By providing a truly interactive conversational environment CPS gives the user capabilities which could never be achieved in a strict batch mode of operation.

Honors and Awards: None

Publications: None

Serial No. 2.15
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Biological Image Processing

Previous Serial Number: Same

Principal Investigator: S. D. Bryan, R. J. Feldmann

Other Investigators: Dr. L. Lipkin, R. Kirsch, Dr. H. Shapiro

Cooperating Units: NIH-NINDS, NBS

Man Years

Total: 1.6
Professional: 1.6
Others: 0

Project Description:

Objectives:

1. The development of methods for man-machine interaction which will permit the trained biologist to represent to the computer the relationships between components of biological images.

2. The implementation of syntax-directed picture processing techniques.

3. The design and development of algorithms for the efficient processing of digitized images.

4. The design of large software systems on the PDP-10 that permit combinations of the programming languages FORTRAN, MACRO and LISP.

Methods Employed:

Continued use has been made of a large LISP program which takes as input a BNF grammar description and a

string from the associated language and produces as output the parse of the string. Preliminary neuromorphological grammars have been written in collaboration with Dr. Lewis Lipkin of NINDS.

Greater emphasis has been placed during the past year on the problem of producing the symbolic input to the grammar, that is, on the problem of going from the raw input data consisting of a digitized image to a determination of the names of the component parts of the image. In this connection, a "morphological analysis" program has been written on the PDP-10 in collaboration with Mr. R. Kirsch of NBS, who originally conceived of the underlying algorithm. The analyzer produces a taxonomy-like tree of the regions of the analyzed picture, based on the adjacency of regions of homogeneous values resulting from spatially differentiating the original image. The analyzer provides not only a means for decomposing the image into syntatic primitives, but a means for determining the degree to which original image information may be discarded when its representation is changed to a collection of regions of homogeneous value with respect to spatial differentiation. (See related reference 1.)

A less research-oriented project involving cytochemistry image data was also conducted. Two PDP-10 programs were developed to process the related photometric data. The first reads digitized scans one at a time from tape, and displays them on the 340 display. The experimenter then uses the Rand Tablet to circumscribe an area on the scan field for which an optical density value is desired. The boundary of this object is displayed. The selection area is measured by counting the number of scan points within the boundary, and a measure of relative optical density is obtained by taking the sum of natural logorithms of the positive digitized video signal values at all points within the boundary. The program allows the experimenter to vary the threshold (i.e., subtract a positive number from the signal value at each point) and to re-draw the boundary of an area of interest until it is satisfactory. Finally, it computes the means and variances of the area measure and optical density measure for all cells on the input tape.

The second photometric processing program also reads single scans. It identifies the largest connected region of the picture after all points below some experimenter-specified threshold value have been removed. This program replaces the manual extraction of an object as described above with an automatic extraction process.

Similar measurements are made on the resulting object. This project is described in detail in the first publication listed with this report.

In conjunction with the morphological analyzer program and the photometric programs, an improved "blob extractor" was written which is believed to be near optimal for our computer and data. The extraction algorithm has been submitted for publication--listed as the fifth publication with this report.

An important part of these early stages of designing an image processing system involves simply determining the quality of the digitized data provided by the scanner, described in related publication (2). For this quality determination we have made extensive use of the PDP-10 340 display, which can display eight distinct gray levels and hence, fairly verisimilar pictures. A program was written which allows the experimenter to adjust in a very flexible way the mapping from the 0-255 original image intensity range to the display's 0-7 range. Scaling of both the original data and the displayed data is also made possible. This program has proved to be very helpful to the biologist user. It is described in detail in publication (2).

Improvements have been made to the PAX II image processing package, the bulk of which was developed at the University of Maryland.

Computer implementation of image processing systems tend to lead to complex combinations of a large number of programs which in general may be written in a variety of languages--in our case the languages being FORTRAN, MACRO and LISP. Two significant contributions have been made toward the relatively straightforward creation of such systems on the PDP-10. One involves the use of a routine that allows jobs to be chained together in a sequence of arbitrary length. The method is superior for our purposes to the overlay mechanism provided by the manufacturer. This system is described in publication (3).

Another routine was developed which permits a single job to consist of LISP-callable FORTRAN and MACRO programs. This allows the combination of the conversation mode of LISP with the superior computational and array features of FORTRAN. In particular, it will allow the creation of a conversational PAX package. This LISP/FORTRAN-MACRO interface program and the chaining routine are being submitted to the Digial Equipment User Society Library.

Major Findings:

The morphological analyzer seems to be the needed link between the raw digitized data and the programs that do the synthetic analysis.

It is clear now from the photometric project, which involved a sample of moderate size, that it is possible already to use the PDP-10 image processing routines for a realistic problem.

The improved blob extractor has proved to be fast enough so that images of moderate size (256 x 256) may be processed in a conversational mode with acceptable response times.

The new machinery for combining LISP, FORTRAN and MACRO routines in a single job and the facility for automatically chaining an arbitrary number of jobs has allowed the design of potentially very large and powerful software systems.

Significance:

The photometric project in itself provides proof that computers may process biomedical image data to provide measurements of accuracy and regional specificity that is all but impossible with conventional methods.

The work with grammars as the means for specifying a large and complex collection of objects, e.g., well formed neuromorphological objects, has been encouraging. Such grammars can be expected to play an important role in the communication between biologists and computers, since they provide a concise representation, whose theory is rather well understood, and peripherally, they offer a basis for the standardization of biomedical descriptions.

The display and Rand Tablet has proved to be very useful in writing programs that the non-programmer experimenter may run with a minimum of training.

There have been a number of software spin-offs that have benefitted other projects, such as the job chaining and language interfacing programs, a magnetic tape handling package, a random-access I/O package, etc.

Proposed Course of Project:

1. The morphological analyzer will be re-written to handle larger images.

2. Further experiments will be performed to test the viability of a large image processing program with mostly FORTRAN routines controlled by LISP.

3. More software will be developed to control the scanner microscope on-line and to process the resulting image data.

Honors and Awards: None

Publications:

1. Shapiro, H. M., Bryan, S. D., Lipkin, L. E., Stein, P. G., and Lemkin, P. F.: Computer-Aided Microspectrophotometry of Biological Specimens. Journal of Experimental Cell Biology.

2. Feldmann, R. J., and Bryan, S. D.: The Representation and Manipulation of Gray Value Raster Picture Data on the Digital 340 Display. In DECUS Proceedings, Fall, 1969, pp. 129-134.

3. Feldmann, R. J.: An Extension to CCL (Concise Command Language). DECUS Proceedings, Fall, 1969, pp. 203-209.

4. Feldmann, R. J., Bryan, S. D.: A Biological Image Processing System. Picture Processing and Psycho-Pictorics, edited by B. Lipkin and A. Rosenfeld.

5. Bryan, S. E.: Contibuted Algorithms. Picture Processing and Psycho-Pictorics, edited by B. Lipkin and A. Rosenfeld.

A. The Strip Blobber: A Blob-Counting and Extracting Algorithm.

B. PAX II FORTRAN Subroutines for Spatial Differentiation and Inverse Distance Transformation.

C. Assembly Language Primitive Routines for PAX II.

Related References:

1. Kirsch, R. A.: Computer Determination of the Constituent Structure of Biological Images, PART 1. National Bureau of Standards Report #10173, December 17, 1969.

2. Stein, P. G. Lipkin, L. E. and Shapiro, H.: SPECTRE II: <u>Science</u>, October 17, 1969.

Serial No. 2.16
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: PDP-10 Systems

Previous Serial Number: Same

Principal Investigator: H. W. Vreenegoor

Other Investigators: H. Lewis

Cooperating Units: None

Man Years

Total: 1/2
Professional: 1/2
Other: 0

Project Description:

Objectives:

1. Development of techniques and devices to permit the use of the PDP-10 computer as a support and back-up for a network of smaller laboratory computers.

2. Development and improvement of the PDP-10 time-sharing monitor.

Methods Employed:

Progress on objective one is reported under the project title, "Communication of Small Computers With the PDP-10 Computer" by Harry Lewis.

Concerning the second objective, a number of new monitor commands were designed and assembled into the PDP-10 monitor providing the capability of linking a network of local and remote Teletypes into one reactive network.

Major Findings:

Experience with the PDP-10 monitor has shown that it is quite feasible to consider adding selected commands to the system without any help from the manufacturer. Also, the design, implementation, and incorporation of drivers for a number of typical devices has proceeded very satisfactorily.

A study of the system under a heavy load of large computer-bound programs mixed with small interactive jobs has shown the desirability of modifying and redesigning the scheduling algorithms and queuing mechanisms in order to improve response time.

Significance:

Continuing improvements to the PDP-10 monitor system will make it easier for the laboratory scientist to communicate data to and from the PDP-10 computer.

Proposed Course of Project:

Continuing development of the monitor with special emphasis on studying and/or implementing techniques of sharing mass storage devices between the IBM 360 and the PDP-10.

Honors and Awards: None

Publications: None

Serial No. 2.17
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Graphics Support for PDP-10 Computer Display System

Previous Serial Number: Same

Principal Investigator: Harry R. Lewis

Other Investigators: M. C. Bruce, H. W. Vreenegoor, S. D. Bryan

Cooperating Units: None

Man Years

Total: 0.5
Professional: 0.5
Others: 0

Project Description:

Objectives:

The goal of the PDP-10 graphics support project is to provide a convenient and efficient man-machine communications system utilizing the graphics-oriented hardware which is part of the PDP-10 computer system.

Methods Employed:

1. We have continued the efforts described in the last report to provide new software to enable users not expert in computer graphics and computer programming to display and manipulate graphic data.

2. New hardware has been designed and built at NIH to further improve the performance of the PDP-10 system in the area of computer graphics.

Major Findings:

1. The compiler-level language for the display has been improved and generalized. The routines in this display library may now be called from LISP, FORTRAN, SAIL, or assembly language. The package now includes routines to accept data from the Rand tablet and function keys, as well as routines to display data on the CRT.

2. A display-tablet comparator has been added to the system. This device simplifies certain graphic operations involving the selection of an item in a display file by means of the Rand tablet. The comparator was designed and built in DCRT.

Significance:

An increasing number of users are making use of the graphic capabilities of the PDP-10 system for exploring a variety of biomedical research areas.

Proposed Course of Project:

Software development to further expand the potential of the PDP-10 graphics system will continue.

Honors and Awards: None

Publications:

1. Lewis, Harry R.: An Interactive Graphics Facility Under the PDP-10/50 Time-sharing Monitor. DECUS Proceedings, Fall 1969. Symposium, Digital Equipment Corporation Users Society.

2. Lewis, Harry R.: Techniques for Generating, Manipulating, and Storage Management of Type 340 Display Files. DECUS Proceedings, Fall, 1969. Symposium, Digial Equipment Corporation Users Society.

Serial No. 2.18
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer Training Program

Previous Serial Number: None

Principal Investigator: Rita G. Minker

Other Investigators: Patricia L. Logan

Cooperating Units: All DCRT Laboratories and Branches

Man Years

Total:	6.75
Professional:	6
Others:	.75

Project Description:

Objectives:

1. To provide the NIH staff with an understanding of and ability to use modern computer technology.

2. To facilitate the formulation and analysis of the problems to be solved so that the technology can be usefully applied.

Methods Employed:

1. Training courses are tailored to the specific hardware and software available at NIH.

2. Seminars focus on both the underlying disciplines (e.g., applied mathematics, computer-related engineering, information sciences) and on the specific areas of potential application.

Significance:

Training courses enable investigators to use DCRT's powerful equipment. Effective use of advanced techniques

and advanced equipment is obtained by further training and directed discussion.

During the past 12 months, over 1200 registrations were processed for the approximately 80 short courses offered.

Proposed Course:

Fall 1970 and Spring 1971 semesters of training classes and seminars are planned.

Honors and Awards: None

Publications:

1. Computer Center Branch: Computer Training Courses--Fall Term 1969. U. S. Dept. of Health, Education and Welfare, Public Health Service, National Institutes of Health, Division of Computer Research and Technology, Computer Center Branch.

2. Computer Center Branch: Computer Training Courses and Seminars--Spring Term 1970. U. S. Dept. of Health, Education and Welfare, Public Health Service, National Institutes of Health, Division of Computer Research and Technology, Computer Center Branch.

Serial No. 2.19
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Molecule Reconstruction from Mass Spectra Data

Previous Serial Number: None

Principal Investigators: R. J. Feldmann, E. Gilbert

Other Investigators: G. W. A. Milne

Cooperating Units: CSL, NHI

Man Years

Total: 0.3
Professional: 0.3
Others: 0

Project Description:

Objectives:

Mass spectra data generated on the high-resolution mass spectrometer of NHI is collected by a small computer. The data is sent to the PDP-10 and it is analyzed to determine possible paths of molecule decomposition. The fragments of decomposition which were determined as molecular formula are converted to a collection of redundant connection tables. The reconstruction of molecules is done in a combinational manner. Some candidate reconstructions are eliminated when the topology fails to meet certain heuristically-determined criteria. The reconstructed molecules are processed to obtain their Wiswesser Line Notation equivalent (WLN). The two-dimensional graph representation of the molecules are then generated from the Wiswesser Line Notation. The programs generated by the project provide a pathway from mass spectra data to graphs of possible molecules. The pathway is a carriage for determining and adding heuristic criteria to constrain the number of reconstructed molecules.

Methods Employed:

The five programs in the system are in FORTRAN. The method of reconstruction has been altered from earlier experiments in LISP so that a FORTRAN implementation is feasible. The Wiswesserization program is the work of A. Chauvenet and C. Farrell. The graph program which uses WLN as input is also being used in a preliminary chemical retrieval program.

Major Findings:

The program components of the reconstruction pathway are in the final stages of debugging. The only conclusion which can be drawn from the programming effort itself is that FORTRAN can be successfully used for topologically-oriented problems. That process requires about 1.5 minutes of PDP-10 CPU time for a 400-weight molecule with 50 mass spectra state.

Significance:

The ability to collect data on a small computer bound to some process must be followed by real-time analysis. The results of data analysis must be presented in a form that is understandable and useful.

Proposed Course:

Heuristic criteria will be added to constrain the reconstruction process.

Honors and Awards: None

Publications: None

Serial No. 2.20
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: PDP-10/SEL 810A Interaction

Previous Serial Number: None

Principal Investigators: R. J. Feldmann, J. DeLeo, S. Allen

Other Investigators:

Cooperating Units: CSL

Man Years

Total: 0.4
Professional: 0.4
Others: 0

Project Description:

Objectives:

First, to provide system-level programming support for the MARC (Medical Audio Response Computer) which is being implemented on the SEL 810A. Second, to provide user-level feasibility analysis and program implementation support.

Methods Employed:

An assembler for the SEL 180A (originally written by W. White in PL/1 on System 360) was converted to PDP-10 FORTRAN. The output from the assembler is passed from the PDP-10 to the SEL 810A via the hardware interface. Additional features were implemented to permit dumping the SEL core image on the PDP-10 line printer and tracing the SEL actions by producing a listing on the PDP line printer. When fully operational, the MARC system will require interactive computational capability. The ability of the PDP-10 to supply this capability is significantly influenced by structuring of user-level programs and data files. In order to study data file

structure in particular, a program was written to retrieve pill data based on partial property specification.

Major Findings:

The development of system programs for the MARC system has been speeded by using the assembler on the PDP-10. The source-level program storage, ability to produce listings, the elimination of paper tape as a data medium, and the rapid transfer of data between the PDP-10 and the SEL 810A are contributing factors. The preliminary pill retrieval program has shown that inverted data files, while providing run time efficiencies, are difficult and time-consuming to update. Multi-level sequential data files seem to be of greatest utility. A second version of the pill retrieval program is being implemented.

Significance:

The interaction of the PDP-10 and the SEL 810A at the system- and user-levels is the first example of the strong interaction between small and large computers of different manufacturers.

Proposed Course:

System- and user-level support for the MARC system will be continued.

Honors and Awards: None

Publications: None

Serial No. 2.21
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Developmental Hardware for the PDP-10 Computer System

Previous Serial Number: None

Principal Investigator: Malcolm C. Bruce

Other Investigators: Harry R. Lewis (comparator project)

Cooperating Units: None

Man Years:

Total: 0.3
Professional: 0.3
Other: 0

Project Description:

Objectives:

This project deals with developmental hardware for the PDP-10 computer system and may include interfacing to other computer systems, non-standard I/O equipment and specialized devices.

Methods Employed:

The design and construction of a hardware Rand tablet display comparator has been completed. The comparator is essentially a light-pen for the Rand tablet and will detect when a point on the display screen is within an electronic square which surrounds the Rand tablet stylus. The PDP-10 system software for the comparator is discussed under project 2.17.

Major Findings:

The comparator allows the Rand tablet stylus to furnish pointing information directly, without the need of extensive software overhead. It has been completed and is just starting to be used in direct application.

Significance:

This is the first in-house hardware to be added to the PDP-10 computer and has demonstrated that the PDP-10 system can be interfaced and modified by DCRT personnel.

Proposed Course of Project:

This is an open-ended project and tentative plans are to take an active role in special data communication hardware for the PDP-10 computer.

Honors and Awards: None

Publications:

1. Lewis, Harry R. and Bruce, Malcolm C.: A Device to Make a Rand Tablet Act Like a Light Pen. DECUS Proceedings, Spring, 1970.

Serial No. 2.22
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1970

Project Title: Computer Utilization Study for the PDP-10 Time-Sharing System

Previous Serial Number: None

Principal Investigators: H. W. Vreenegoor, S. D. Bryan, R. J. Grunby

Other Investigators: None

Cooperating Units:

Man Years

Total: 1
Professional: 1
Other: 0

Project Description:

Objective:

To produce reports which show how individuals are using the PDP-10 resources and what demands are made on the system as a whole. This information may be used for billing customers, for helping individuals improve their use of the system, and for identifying areas where system change may improve service.

Methods Employed:

Program 1 (called SNOOPY) captures system information once every five minutes, such as number of users sharing the system, how much CPU time, core, disk space, etc., is used by each, and what programs are running. This SNOOPY data is collected on the disk.

Program 2 daily flushes SNOOPY data from disk onto some convenient external device, normally Dectape.

Program 3 will normally be run once a month to produce on magnetic tape a master file of all SNOOPY data collected to date.

Program 4 and 5 can look at SNOOPY data within any selected time range and produce a summary report or a detailed report of computer usage on the printer. This report makes extensive use of histograms to illustrate how much load is on system resources, such as core, disk, Dectape, utility programs, and use of time-sharing terminals; and also how individuals compare in their use of these resources.

Proposed Course:

Completion of current plans.

Honors and Awards: None

Publications: None

Serial No. 2.23
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: On-line File Processing System

Previous Serial Number: None

Principal Investigator: Jon A. Halverson

Other Investigators: Marvin Katz, Richard Baxter

Cooperating Units: None

Man Years

Total:	2-1/2
Professional:	2-1/2
Others:	0

Project Description:

Objectives:

The primary objective is to establish an on-line file processing capability at the NIH whereby the user can maintain, retrieve and display information in a data base via a remote terminal.

Methods Employed:

The system operates on any data base whose organization is indexed sequential. A macro language is provided which permits the terminal user to write application programs on a data format and functional level, thus reducing substantially the programming effort and time required for installation.

Significance to Program of the Division:

The system can be considered as an initial step toward the development of on-line information systems for the NIH. For the first time investigators responsible for maintaining large data bases will have a powerful means of accessing critical information quickly.

Honors and Awards: None

Publications: None

Serial No. 2.24
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Commuication of Small Computers with the PDP-10 Computer

Previous Serial Number: None

Principal Investigator: Harry R. Lewis

Other Investigators: None

Cooperating Units: NIND, DCRT-CSL

Man Years

Total:	0.25
Professional:	0.25
Other:	0

Project Description:

Objectives:

To meet the needs of various small computers on the NIH campus for computing back-up and mass storage on the PDP-10 computer system.

Methods Employed:

Small computers are linked to the PDP-10 by two methods: 1. Over ordinary telephone lines, using high-speed data sets at each end and encoding the data transmitted as Teletype messages received and dispatched by the PDP-10 Data Line Scanner. 2. Over a special interface connecting the I/O bus of the PDP-10 with that of the satellite computer.

Major Findings:

1. Communications with a PDP-12 and a LINC-8 in Building 36 have been established over telephone lines using

1200-baud data sets. This type of transmission is the easiest to accomplish from the standpoints of both hardware and software, and is adequate when the amount of data being transmitted is relatively small.

2. The PDP-10 is communicating with the SEL 810b computer in the basement of Building 12A over a specially-designed I/O-bus-to-I/O-bus interface. This technique yields a much higher data transmission rate. The module of the PDP-10 monitor which services this interface was written so as to be re-entrant. Hence, if other small computers are linked to the PDP-10 computer using this type of interface, no new software will need to be developed to accommodate them on the PDP-10.

Significance:

Increasing numbers of small laboratory computers are being acquired by NIH.

The experience gained with these efforts toward computer-to-computer communication will expedite the implementation of further interfaces. Already the PDP-10 is expanding the usefulness of small computers by enabling them to utilize the PDP-10 peripherals, such as mass storage media and the line printer, and by providing computational power not previously available.

Proposed Course of Project:

We will continue to study the needs of small computers for access to a large computer and to develop ways to meet these needs.

Honors and Awards: None

Publications: None

Serial No. 2.25
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Optical Document Processing System

Previous Serial Number: None

Principal Investigator: Elliott Alterman

Other Investigators: None

Cooperating Units: None

Man Years

Total:	.5
Professional:	.5
Other:	0

Project Description:

Objectives:

NIH processes vast volumes of data in many forms and formats. The effort necessary to analyze and process the data currently requires significant time and effort to convert these data to machine readable form. We are attempting to eliminate much of this intermediate data conversion by using the documents themselves as input to the central facility computers.

Methods Employed:

Internal processing routines are being developed to process documents read by the optical character recognition equipment (IBM 1288 Optical Page Reader). Evaluations and discussions of areas of possible application and their unique problems and requirements, operational procedures and reliability and performance considerations are currently underway. Included as a goal of the project is a unique attempt to have all necessary user-defined instructions for the processing of the documents be contained on one of the documents to be read by the optical character recognition equipment.

Significance to Program of the Division:

The succesful completion of this project will enable CCB,CR to more fully meet the data processing needs at NIH in a flexible and efficient manner. The current cost of many of the data processing activities will be reduced significantly and the design goals of efficient and reliable performance will enable CCB,CR to perform other activities currently deemed too expensive or time consuming for practical consideration.

Honors and Awards: None

Publications: None

Serial No. 2.26
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Extensible Language Project

Previous Serial Number: None

Principal Investigator: Jon Halverson

Other Investigators: P. Kalkowski, M. Katz, G. Knott

Cooperating Units: None

Man Years

Total: 1
Professional: 1
Other: 0

Project Description:

Objectives:

The purpose of the project is to provide a programming language which can be extended to include new syntactic constructions.

Methods Employed:

A base language is defined which contains primitive data types, operations, and statements. The syntax of the base language is specified in BNF notation and encoded in syntax tables. An analyzer program, given these syntax tables, will generate a parse for a statement written in the base language. A compiler for the base language will invoke the analyzer, process program text parsed and transformed by the analyzer, and produce a compiled program.

Extensions to the base language can be implemented provided that the extensions can be expressed in terms of the primitive constructions of the base language.

The syntax for the extensions is added to the syntax for the base language, new syntax tables are produced, and programs are written to transform constructions in the extended language to constructions in the base language. These transformation programs are invoked by the analyzer upon recognition of the extended language constructions during parsing.

Significance to Program of the Division:

An extensible language facility reduces the work involved in generating language processors and makes possible the definition of different languages to handle different kinds of problems or to handle problems which require special notation. Moreover, the problem-oriented languages so defined include just those constructions which are relevant to the problems under consideration.

Honors and Awards: None

Publications: None

July 1, 1969 through June 30, 1970

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT - 3

2. LABORATORY OF APPLIED STUDIES — 3. Eugene K. Harris, Chief

I. GENERAL SUMMARY

The Laboratory of Applied Studies engages primarily in cooperative studies, with NIH scientists, applying mathematics, statistics and computing science to biomedical problems. To support these studies, special-purpose computer programming systems are developed where needed; these are often later extended for more general use and distributed to potential users in the form of DCRT technical reports.

In addition, the Laboratory undertakes self-generated mathematical and mathematical statistical research with applications to biological studies. More recently, LAS has assumed responsibility for the guidance of statistical programming services required in the analysis of large-volume multivariate studies.

This year's report has been substantially simplified and shortened in an attempt to improve communication and reflect the coordinated nature of many individual projects. As a result, only five project reports are included, each covering a number of "sub-projects" contributing to the same general objective.

II. CURRENT LABORATORY PROGRAMS

1. Cardiovascular and other Collaborative Studies

In previous years, the Laboratory has devoted major efforts to projects with NHLI and other (university) scientists on cardiovascular dynamics using data acquisition and programming systems specially designed for hemodynamic data. These projects have included, for example, study of energy losses across artificial valves and hemodynamic changes observed during cardiac surgery and in the cardiac catheterization laboratory. While such investigations are continuing, (Project Report 3.3), LAS has this year initiated projects related more closely to diagnostic and medical care facilities in the area of heart disease (Project Report 3.2). Specifically, LAS has begun evaluation and comparison of several computerized systems of ECG analysis, and is attempting to describe quantitatively the operation of coronary care units. Such projects partake of the methods of operations research with emphasis on the purposes and nature of the system under study, the construction of descriptive mathematical models (both deterministic and probabilistic), computer simulation and identification of criteria for measuring costs and benefits.

The essential LAS contribution to cardiovascular studies is the multidisciplinary team approach. The LAS group includes a board-qualified internist, electronic engineers, mathematicians and statisticians, computer specialists and programmers, drawn from all sections and units of the Laboratory and able to obtain immediate help if needed from the Computer Systems Laboratory and other DCRT labs and branches. Individual special talents are pooled to gain a greater understanding of the biomedical problem as well as to test and implement the necessary computing methods. The association with the NIH scientist(s) then becomes a true cooperation in the performance of research, not simply a consulting or service function.

The same spirit of close cooperation in research naturally guides other collaborative studies where only one LAS staff member is working with one scientist from another Institute. Two such studies, are a) the mathematical modeling of neuronal response to synaptic input through dendritic spines and b) studies on the protein binding of fatty acids and thyroxine (Project Report 3.5).

Continuing analysis of clinical chemistry data, in cooperation with the Clinical Pathology Department, NIH, has produced a series of reports now in press. (Project Report 3.4) Results to date have shown that, although very few blood constituents are sufficiently powerful by themselves to clearly distinguish one normal blood profile from another, substantial variations in average blood levels of many constituents exist even among normals of the same age-class, sex and race. Continued improvement in analytic methods is needed to prevent obscuring intra-individual biological variation.

2. Statistical Research; Statistical Programming Activities

During the past year, continued progress has been made on fundamental statistical theories of size and shape, a basic research interest of the Biomathematics and Statistics Section with potential applications to human growth and differentiation (Project Report 3.1). Associated with these studies, a new and extremely general technique has been worked out for classification of observations into an unspecified number of homogeneous clusters. Immediate application of this method will be to the study of normal variations in blood chemistries, mentioned above.

Among the many activities of the Statistical Programming Unit of LAS, attention is drawn here to three in particular: 1) development of programs for analysis of variance of repeated observations (not necessarily independent) of multiple factors, 2) design of a storage, retrieval and analysis system for blood chemistry data, 3) testing and updating of statistical library programs together with general assistance to NIH users.

3. Education and Training

During 1969-1970, Laboratory staff conducted courses for NIH scientists in the areas of time series analysis, differential equations, multivariate analysis, graph theory, complex variable theory, use of the CalComp Plotter, elementary probability for biologists, regression and correlation and others.

4. Establishment of Applied Mathematics Unit (see Project Report No. 3.5)

An Applied Mathematics Unit, headed by Mr. John E. Fletcher, has been formed to consult and collaborate with NIH scientists on mathematical aspects of their research, and to do whatever programming is needed to support this work. Although a direct descendant of the earlier Mathematical Programming Unit, the new group will emphasize those projects offering extended collaboration with NIH research leaders interested in the applications of mathematics and computers to biomedical problems.

5. Purchase of Laboratory Computer

Past experience in LAS-I/D collaborative projects has shown that such work would benefit greatly if a portable, fast computer with analog-digital conversion capabilities were available during the experimental procedures. After reviewing bids from a dozen companies, delivery was taken on a MAC-16 (16-bit word, 8k-storage, 1 μsec cycle time, with CRT and magnetic tape drive), now undergoing acceptance tests. Its use is planned in both cardiovascular and neurological research projects.

LAS reports published or "in press" during FY 70

1. *Connor, R.J.: The sampling distribution of the range test for homogeneity J. American Statistical Association. 64: 1443-1471 1969.

2. Spector, A.A., John, K., and Fletcher, J.: Binding of long-chain fatty acids to bovine serum albumin. J. of Lipid Research. 10: 56-67 1969.

3. Spector, A.A. and Fletcher, J.: Fatty acid binding to beta-lactoglobulin. Lipids J. (in press)

4. Gilbert, D.B., Nolan, S.P., Stewart, S., Fogarty, T., and Harris, E.K.: An in vivo study of energy losses of ball and disc valves in the mitral position. J. of Applied Physiology. 28: 282-290 1970.

5. Harris, E.K.: Distinguishing physiologic variation from analytic variation. J. of Chronic Diseases. (in press)

6. Harris, E.K., Kanofsky, P., Shakarji, G., Cotlove, E.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, II. Estimating biological components of variation American J. of Clinical Pathology. (in press)

7. Cotlove, E., Harris, E.K., Williams, E.Z.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, III. Physiological and medical implications. Amer. J. of Clinical Pathology. (in press)

8. *Hutchinson, G.: Evaluation of polymer sequence data using graph theory. Bull. of Math. Biophys. 31: 1969.

9. Mosimann, J.E.: Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. J. of the American Statistical Association. (in press)

10. Mosimann, J.E.: Discrete distribution models arising in pollen studies, symposium random counts in scientific work. Penn State Univ. Press. (in press)

11. Mosimann, J.E. and Greenstreet, R.: Representation insensitive methods for paleoecological pollen studies. Penn State University Press. (in press)

12. Minicozzi, W.P. and Stroot, M.T.: On the determination of interaction energy functions. J. of Computational Physics. (in press)

13. Zaveler, S. and Ashbrook, J.D.: CalComp manual, DCRT Technical Report No. 3.

14. Mejia, R.A. and Chang, C.: Time series analysis, DCRT Technical Report No. 4.

*paper cited in last year's annual report as being "in press"

Serial No. DCRT 3.1
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Biomathematics and Statistics

Previous Serial Number: 3.1, 3.2, 3.11, 3.13, 3.20

Principal Investigator: J. E. Mosimann
assisted by: G. Hutchinson, R. Greenstreet, G. Atta, S. Selkow

Co-Investigators: P. Berthet, Univ. of Louvain, Belgium
P. Jolicoeur, Univ. of Montreal, Canada
C. Merril, LGCB, NIMH

Objectives:

1) To conduct mathematical statistical research in areas of biological importance, particularly in a class of problems designated "size and shape".
2) To develop computing methodology of value to biostatisticians and other health scientists.
3) To investigate the application of mathematical theory to molecular biology and biochemistry.

Progress During Past Year:

a) Size, Shape and Pattern Statistics:

The goal of this continuing project is to provide a solid foundation in statistical theory for the study of size-associated shape variables in organisms, including man.

Size and shape variables are first defined within a probabilistic framework. For example in man the ratio head-length/trunk-length is a shape variable. Trunk-length, head-length, and (head-length + trunk-length) are three examples of size variables. Since human infants have relatively larger heads than adults, the shape variable head-length/trunk-length changes with size (and age) in man. Previous studies of size-related shape changes have used functional equations like

$$y = ax^b$$

to describe these phenomena. These studies have not explicitly chosen a size variable to which shape is related. The probabilistic framework reveals that this choice is important and is implicitly

made in all studies. One result is that shape can be independent of at most one size variable.

Related to these studies, a test of significance has been constructed for the independence of size and shape variables distributed in the multivariate lognormal form. Currently, the distributional properties of a "shape" statistic, Shannon's index of diversity, are under study. An earlier study of the effects of unequal scale changes of variables on shape statistics (proportions and ratios) has been completed and is in press (see LAS publications list).

b) Computer Techniques for Clustering Analysis:

As part of this study, existing clustering techniques, invariant under unequal linear scale changes, were investigated, and a new pattern recognition or clustering method developed. This new technique presupposes no prior knowledge of the distributions of the variables studies, nor any specified number of clusters. The system uses a potential function as a measure of similarity to determine which observations to group together. The method may also be used to classify new observations. A battery of clustering programs are now available on the 360 computing system to all interested NIH users.

c) Mathematical and Computing Methods in Molecular Biology and Biochemistry:

A recent paper on the application of algebra and graph theory to nucleic acid sequencing (Hutchinson, 1969 - See LAS publications list) led to the author's gaining firsthand experience in standard laboratory techniques in a biochemistry laboratory (Lab. of General and Comparative Biochemistry , courtesy of Dr. Carl Merril, NIMH). Experiments were designed and conducted to determine DNA sequence information by means of RNA-dependent DNA polymerase reaction. Although definitive results were not obtained, the ensemble of these laboratory experiences greatly increased Dr. Hutchinson's understanding and ability as a mathematical collaborator with research workers in molecular biology. As one result of these efforts, he has become engaged in a cooperative project to study through a simulation program on the PDP-10, the logical sequence of steps which comprise genetic control mechanisms in bacteriophage.

Serial No. DCRT 3.2
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Evaluation of Health Care Systems

Previous Serial Numbers: None

Principal Investigator: J. J. Bailey, M.D., assisted by M. Horton, R. Greenfield

Co-Investigators: R. Smith, Mayo Clinic
D. Beiser, NHLI

Project Description:

Objectives:

Two sub-projects are currently being conducted in the general area of the evaluation of health care systems. These are, a) evaluation of computer diagnosis of electrocardiograms (ECG), b) Cost-effectiveness study of coronary care units.

a) LAS plans to test the validity of the Mayo Clinic Program for automated ECG analysis in cooperation with cardiologists at NHLI. Further, we hope to show that a cost saving accrues when the cardiologist has the computer output available as he reads the EKG. Recent studies have suggested that ECG reading may be reduced by a factor of 5-to 10- fold with consequent dollar savings in cardiologists' time. It is also possible that in many instances the computer would include statements neglected by humans. If so, an estimate of upgrading due to computer processing should be included in a cost-effectiveness study. Finally, cost-effectiveness is likely to be sensitive to patient distribution. The system may show a pay-off in a population which is largely normal but may not in a population with many abnormal and bizarre cases.

We hope to bring to NIH for comparison and evaluation other ECG analysis programs such as those developed at Mt. Sinai (L. Pordy) (Medical Systems Development Lab (C. Caceres) and the V.A. (H. Pipberger). Ultimately one ECG program should evolve from all these efforts. Comparison and cost-effectiveness studies have great interest for physicians and institutions not only across the United States but also in Western Europe.

b) The Coronary Care Unit (CCU) represents a health care subsystem which may be amenable to a systems analysis approach. A need exists to determine optimal bed capacities for communities of different sizes and to develop a methodology for estimating the net costs and benefits accruing from CCU operations.

Analytic techniques developed for the CCU will be directly applicable to other health care subsystems such as surgical recovery units, special diagnostic wards, respiratory care units, stroke wards, etc.

The Veterans Administration Task Force IV on Delivery of Health Services, the Interregional Management Information System (for Maine, Vermont, New Hampshire, Massachusetts and Rhode Island), and numerous hospitals have expressed interest in a cost-effectiveness model, especially with validation. An abundance of data appears to be available from local and out-of-state hospital sources.

Current Status:

a) Automated ECG diagnosis: The Mayo Clinic program is now running on a 360 system. A certified 360 version of the MSDL program is expected to be completed and delivered to us shortly.

A protocol for a comparative study of manual vs. computer analysis of ECG's is now being drafted by LAS in collaboration with Dr. Beiser of NHLI. This project is supported by the data transmission and analog to digital conversion system initiated by LAS during FY 1969 (Project Report No. 3.17) and completed during this reporting year. This system permits the acquisition of analog ECG data on the clinical ward, conversion to digital form with a time/event code reader and processing on the IBM-360 through any of the several ECG analysis programs.

b) A deterministic cost-effectiveness model has been designed to relate the principal CCU parameters: mean arrival rate, mean service time, admissions policy, infarct fraction, partitioning of patient population, and cost analysis. A computer program to simulate the stochastic aspects (related to queuing theory and involving inter-arrival and service-time distributions as well as specified bed capacity) has been constructed. Data for validation will be soon available from several hospitals.

Serial No. DCRT 3.3
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer Methods in Cardiologic Research

Previous Serial Number: 3.5, 3.6

Principal Investigator: J. J. Bailey, M.D.
assisted by: M. Horton, M. Douglas, W. Scott

Co-Investigators: S. Epstein, S. Levitsky, W. Whitehouse (CB, NHLI)
J. Talano, Georgetown University

Project Description:

A. Cardiovascular Dynamics

Background and Objectives:

For the past four years, LAS, in cooperation with the Cardiology Branch and Heart Surgery Branch, NHLI, has continued developing and applying computer-based methods for determining the dynamic cardiovascular status of patients on a beat-by-beat basis. These efforts have produced an extensive acquisition and computer processing system for hemodynamic studies (DCRT Technical Report No. 2), with application to research projects, e.g., heart valve studies (Gilbert, Nolan et al. - see LAS publication list) and care of heart surgical patients.

An important new effort in this direction concerns the automated determination of instantaneous ventricular volumes by angiograms. When these data are combined with other information from the cardiovascular catheter laboratory, beat-to-beat estimates of flow, work, power, ventricular contractility, and ventricular compliance can be measured on patients with a variety of conditions including valvular problems and coronary vessel disease. This should lead to better understanding of pathophysiology as well as more accurate diagnosis and prognosis of cardiac patients.

Current Status:

Hemodynamic programs are being modified to calculate maximum rate of pressure rise (an index of ventricular contractility) for use in the study of patients with Idiopathic Hypertrophic Subaortic Stenosis. In each of these patients, the Cardiology Branch, NHLI, anticipates collecting 10-20 minutes of data. A population of some 20,000 beats

will be available for analysis through these hemodynamic programs. A considerable amount of professional time could be saved and the efforts of the catheterization laboratory maximized if these results were immediately available. Moreover, the cardiologist could then decide while the patient was still in the laboratory whether additional studies were needed. To this purpose, appropriate parts of the present LAS hemodynamic programming system plus supplemental 360 programs recently developed are being implemented on the LAS MAC-16 computer.

Angiographic data, processed by the TV laboratory in Building 10, has been shown to correlate well with cardiologists' measurements. At this time, these data are being formatted with other data for input to the hemodynamic programs.

The Cardiology Branch, NHLI, is arranging to purchase an acquisition device recommended by LAS to enable gradient computer processing of catheterization laboratory data.

B. Analysis of Vectorcardiograms

Background and Objectives:

Specific motivation for this project has come from cardiologists at NIH and Georgetown University interested in the sequential study of vector loops following myocardial infarction. In addition there is general interest in the so-called ventricular gradient, an index of differences between depolarization and repolarization pathways. The standard approach has been to measure the angle between the mean QRS vector and the mean T vector. An angle of more than 90 degrees is considered abnormal. A reasonable alternative would be to measure the angle formed by the normal vectors to the best fitting planes of the QRS and T loops. Another approach suggested by Abildskov is to construct the so-called primary T wave. The primary T wave is the difference between the observed T wave and the "secondary T wave". The latter is calculated from the QRS pattern (depolarization) and associated action potentials.

Current Status:

Several versions of a vector loop program on the CalComp machine have been completed. Further, programs to find the best-fitting planes for QRS and T waves, and the angle between the normals to these planes, have been finished and await testing. An algorithm to simulate the Abildskov primary wave is currently under development.

C. Computer Modeling of Cardiac Conduction

Background and Objectives:

The objectives and methodology of this continuing project were stated in the Annual Report for FY, 1969 (Project Report No. 3.6). The primary goal is a realistic model for simulation and, hopefully, better understanding of both normal heart function and pathologic conditions such as infarcts, conduction

disturbances, ventricular aneurysms, etc. The gross anatomy of the heart is represented by small building blocks or cubes of "myocardium". The locations of cubes are specified within four pairs of concentric ellipsoids, the heart chambers. Each cube is assigned a tissue type, viz. non-conducting, atrial tissue, A-V node, Purkinje cells, or ventricular muscle. Each tissue type has a characteristic action potential and velocity of conduction. A "stimulus" at the sino-atrial node is then propagated throughout the heart. The time at which a cell (cube) fires is a function of its neighbors' firing time and velocities of conduction. The transmembrane potential of a cell depends upon its action potential, a stored template. This electrogram becomes the surface electrogram when appropriate transfer impedances are applied. This surface electrogram can be transformed into a 12 ECG display by a 12-lead simulator.

Current Status:

The geometric model of gross anatomy is complete and has been displayed on the DCRT adage computer system. The ECG simulation requires adjustments in the spacing of the Purkinje fibers before the correct depolarization pathways are realized and before further development can proceed. Construction of the ventricular dynamic model is underway.

Serial No. DCRT 3.4
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Statistical Research in Clinical Pathology

Previous Serial Number: Same

Principal Investigator: Eugene K. Harris, Ph.D.

Co-Investigators: Ernest Cotlove, M.D., Clinical Pathology Dept., CC
Donald Young, M.D., Clinical Pathology Dept., CC
George Shakarji, LAS, DCRT

Background:

Previous annual reports have amply described the nature and objectives of this continuing project. Briefly, efforts during the past several years have been devoted to the study of analytic and biological components of variance in normal blood chemistries. The data base has consisted of 68 screened normal subjects, each providing 10-12 weekly blood samples, collected under controlled conditions and analyzed in duplicate for 15 common constituents. Source data for estimating long-term analytic variance have been provided by daily serum pool analyses during a two-year period.

Current Status:

The first phase of this study has been completed. Using standard statistical methods, analytic and biological variance components in blood chemistry tests have been estimated within and among normal individuals. A series of three reports on this work, with implications for laboratory analytic methods as well as interpretation of normal ranges has been completed and awaits publication (see LAS publications list). These studies have shown (a) that where blood constituents under tight homeostatic regulation are concerned (e.g. sodium, magnesium, calcium), even the best current methods of analysis often mask small biological changes occurring in periodic controlled samples from a normal individual; (b) although very few of the tests studied (uric acid, cholesterol are exceptions) appear sufficiently powerful by themselves to distinguish among normals (i.e. to offer individual blood profiles), significant differences did appear among normal individuals with respect to mean levels in all constituents studies. Such differences were generally not explainable on grounds of age, sex or race, casting doubt on the general usefulness of "normal ranges" specific for these variables; (c) need exists for better methods of distinguishing biological from

analytic variation within a single individual. Some statistical research on this problem has been completed, and a non-parametric method developed for estimating the mean and variance of a distribution of intra-individual biological variances. A separate report on this topic is now in press.

An experiment to uncover intra-individual biological variance by eliminating long-term analytic variance has been conducted and results are now undergoing analysis. In this experiment weekly samples are stored in frozen state and all analyses performed in one day at the end of the collection period. Plans for FY 1971 include application of multivariate statistical methods to determine relationships among blood tests. It is hoped that these analyses, and those recently completed, will aid in the evaluation of blood chemistry results in routine medical diagnosis, including multiphasic health screening.

Serial No. DCRT 3.5
1. Laboratory of Applied Studies
2. Applied Mathematics
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Applied Mathematics

Previous Serial Number: None

Principal Investigator: J. Fletcher, Head Applied Mathematics Unit
assisted by J. Ashbrook, M. Douglas, E. Hill, J. Rinzel, M. Stroot

Co-Investigators: G. Weiss, PSL, DCRT
W. Rall, NIAMD
R. Mejia, LAS, DCRT
A. Spector, Univ. of Iowa

Background and Objectives:

Established within the Laboratory of Applied Studies during this reporting year, the primary responsibility of the Applied Mathematics Unit is to provide NIH scientists with mathematical competence for biomathematical modeling and data analysis. This competence includes both theoretical and applied techniques, as well as numerical computation methods. Each individual in this unit has a primary specialty in computer science or mathematics, and all are capable computer programmers.

Current collaborative projects involve: modeling of biological systems by ordinary and partial differential equations using both theoretical and numerical methods; fitting non-linear models by optimal procedures; conversion of analog information to digital data; the design and development of interactive computer modeling systems. These are described briefly below:

Some Current Projects:

1. Numerical Inversion of Laplace Transform - Laplace transforms are used to solve linear differential equations arising in the modeling of a physical or biological system. Employing techniques of advanced mathematics and numerical analysis, a general algorithm has been developed for the numerical inversion of Laplace transforms not invertible by other means.

2. Molecular Model Building Using Theoretically and Empirically Determined Intra- and Inter-Molecular Potential Functions - Knowing

the electrostatic fields that molecules experience, it should be possible to understand and solve biological problems on a detailed molecular basis. A general computational procedure has been developed for deriving molecular interaction energy functions from equilibrium crystal phase data. Current investigations center on the search for a single model which would explain the empirical data of several classes of crystals.

3. On-Line Modeling System - An interactive computer display system, designed and implemented on the PDP-10 computer, allows users to specify mathematical models, manipulate data files, change parameter values and view results in real time. The teletype, Rand tablet and function keys are used as input devices. The system has been tested on neural network programs and used to study such stimulus-response relations as those embodied in the Hodgkin-Huxley neural conductance equations. A DCRT Technical Report describing the system is in preparation.

4. Mathematical Modeling of Neural Activity - A mathematical model has been developed to investigate the response of a neuron to synaptic input at a dendritic spine. The relative isolation of exciting inputs from spine synapses as opposed to those from synapses directly on the dendrite surface has been found to depend on such parameters as spine stem resistance. Using the neuron's response function for injection of current to a dendritic branch, the mathematical model was cast as a system of Volterra integral equations. An efficient numerical technique, involving the concept of product integration, was devised to solve this system of equations. The same approach may prove valuable in the analysis of other systems involving partial differential equations for which response functions can be determined. Results are being prepared for presentation and publication (Biophysical Journal).

5. Neuroelectric Data Retrieval - Programs have been completed for the analysis of events, such as neuronal discharges. One such program uses the EAI 8900 hybrid computer system to retrieve and separate (through amplitude windows) spikes from up to fifteen cells recorded on analog instrumentation tape. A complete description of this system is now being prepared.

6. Macromolecule-Ligand Binding - Studies are being made in the following areas: a) Equivalence of binding models, b) The effects of noise in various types of binding data, c) Computational methods of fitting models to data, and d) The resolvability of a model's parameters from a given set of experiemental data. Applications of these areas are being made to the binding of

a) Fatty-Acid to Albumin (HSA)
b) Fatty-Acid-Albumin-Mammalian Cell Transfer
c) Thyroxine to Albumin (BSA)
d) Structure of Albumin Molecule in the Binding Process

7. CalComp - DCRT's Technical Report No. 3: "Programmer's Manual

for CalComp Plotting", prepared by staff of the Applied Mathematics Unit, LAS, has been distributed throughout the NIH scientific community. This manual is intended to serve as a reference and a user-oriented introduction to the CalComp plotter at the DCRT Computer Center. Based on this manual, a training course entitled "Introduction to the CalComp Plotter" is given periodically within the DCRT training series.

Allocation of Staff Time During FY1970

Project		Man-Years		
		Laboratory of Applied Studies		Cooperating
		Professional	Supporting	Investigators
3.1	Biomathematics & Statistics	5	1	.25
3.2	Evaluation of Health Care Systems	2	1.5	.25
3.3	Computer Methods in Cardiologic Research	1	2.0	.5
3.4	Statistical Research in Clinical Pathology	1	1.0	.3
3.5	Applied Mathematics	6	0	1.5
	Statistical Programming	6	1	2
	Administrative	1	.5	
	Totals	22	7	5.3

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT - 4
Serial Number

2. COMPUTER SYSTEMS LABORATORY

3. Alan Demmerle
Chief

The Computer Systems Laboratory (CSL) concentrates on data system problems that do not lend themselves to solution through use of DCRT's central computer facility and require, instead, the development of peripheral computer systems. The mission of CSL is the design and implementation of computer based data systems for any promising application at NIH, even though, during our first few years, we have concentrated on the areas of laboratory automation and the delivery of health care. These areas are approached from a total system point-of-view. A specific area which promises potential benefit by the application of the computer technology is identified; the computer techniques to be used are chosen, the hardware and software aspects of the system are designed and implemented, and the users instructed how to use the system. Such work requires a multidisciplinary group effort and requires two to four years from inception to completion.

Computers in the Laboratory

The laboratory automation work, to date, falls in two categories: the "computerization" of biochemistry laboratories, and the "computerization" of psychological studies. In the first category there is the computer system for NIDR, one for NIAMD, one for the NHLI's Mass Spectrometers, one for the NICHD's Gerontology Research Center in Baltimore, and one for use with the PSL-DCRT NMR Spectrometer. There are some factors which are common to all these systems. They are all used to collect data directly from instruments, to calibrate, to format, to perform calculation on the data, and to produce analysed experimental results for the bench scientist during, or immediately after, his experiment. The processed data is then available in a form which is particularly easy to use by the experimenter, for example, as plotted results or graphical results displayed on a CRT. Generally, the system collects data from several instruments simultaneously so the marginal cost per user, or per experiment, is minimized. The systems provide the scientist with analysed data at the time he is conducting his experiment, thus allowing him to modify his experiment while it is still set up and his experimental samples are still available. It allows him to do some experiments that he could not do before, for example stimulus-response experiments using implanted micro electrodes can be performed, the stimulus changed and more data collected and analysed before the cell in which the electrode is implanted dies. These computers also allow him to analyse more data in more complex ways than was ever possible before.

In the second category is the NIMH (Building 10) system to be used to facilitate psychological studies. In these studies the patients interact with the computer. Here, also, the computer is shared among a number of concurrent studies - each scientist can use the computer simultaneously yet independently, of the others. The system serves to collect data, control and modify game situations and control and format displays of various kinds.

Computers in the Delivery of Health Care

There are three projects in this Laboratory relating to the delivery of health care. The first one is an effort to bring easy to use, economical and reliable computer services to the practicing physician. He will be able to use a ubiquitous terminal such as a standard telephone, to access commercial computer systems for assistance in such activities as diagnosis and therapy planning. This system is designed with the community physician in mind, as well as NIH clinical requirements. The second project relates to the care of critically ill patients. In this project the general purpose computer is used as an aid in the design of more convenient, more reliable, and more economical special purpose hardware for monitoring arrhythmias and cardiac output. The third project is related to the automatic processing of medical English text, and is viewed as the only feasible way to make the vast quantities of textual research data, such as autopsy and surgery reports, available for analysis and the testing of hypothesis.

Most of these projects began prior to this reporting year. Most represent collaborative work with other Institutes. Attached is a table which shows graphically the CSL personnel effort put into these projects over the last two years, the capital expenditure required for these projects and the source of these funds. The capital expenditures portion of the table has one entry for equipment which includes both commercially purchased equipment and the cost of components for the equipment developed in this Laboratory, and one entry for central facility computer charges expended on the project for computer program development. The personnel effort includes that required for problem analysis, system engineering, equipment design engineering and programming, all of which is done by professionals in the field of engineering, mathematics, physics, chemistry, and computer science.

It is our general policy to buy competitively from industry as much equipment and software as is commercially available for these projects, and design and produce here that equipment and software for these systems which is not commercially available.

No additional major projects are planned for the coming year because we anticipate that all of our personnel resources will be required to continue development of these projects. It is anticipated that some of these projects will be complete by the end of FY-71.

During the Fiscal Year 1970, a group of four engineers received a superior accomplishment award for their efforts on behalf of the Myocardial Infarction Research Program of NHLI.

		CSL MANPOWER		CAPITAL			
				Equipment			Computer Charges
Project #	Name	FY-69	FY-70	Source	Amount	FY	
DCRT 4-1	NIDR	5(m-y)	5 (m-y)	NIDR	100(K$)	68	
				DCRT	150	69	
				DCRT	0	70	12(K$)
4-2	NIAMD	5	5	NIAMD	180*	68	
				DCRT	25	69	
				DCRT	20	70	3
4-3	Mass Spec.	2	2	DCRT	76	69	
				DCRT	10	70	½
4-4	Gerontology	1	3	NICHD	111	69	
					0	70	0
4-5	NMR	0	½	DCRT	90	70	0
4-6	NIMH	½	1	NIMH	142*	70	0
4-7	Med. Telecomm.	6	6	DCRT	125	69	
				DCRT	112	70	12
4-8	Cardio-Vascular	3	5	DCRT	20	69	
				DCRT	21	70	1
4-9	Auto Processing of Med. English	4	4	DCRT	0	70	30
4-10	Consultation	2	1½	DCRT	0	70	0

*This is the purchase price of the system which was partially purchased and partially leased.

(Cost figures are approximate.)

<u>List of all papers written and published by CSL members</u>

<u>during FY 1969-70</u>

Brown, Leroy, Brown, Fred, Wilson, John, and Dubner, Ronald: Real-Time Analysis of Neurophysiological Data in a Multiprogramming Environment. (Not submitted yet) <u>Council To Advance Programming</u>, June 7-8, 1970.

Gilbert, Eli J.: Automated Generation of Molecular Structure from Mass Spectrometry Data. Paper No. 138. (Abstract) <u>43rd Fall Meeting of the American Oil Chemists Society</u>, Minneapolis (1969)

Gilbert, Eli J.: Mass Spectrometry Data Processing at NIH. (Chap. in a book) Fall of 1969.

Kempner, Kenneth, Miller, Martin, and Holsinger, William: A Computer Approach to Arrhythmia Monitoring. <u>8th International Conference on Medical and Biological Engineering</u>. Palmer House, Chicago, Ill., July 20-25, 1969.

Otten, Michael, Allen, S. I., Plexico, Perry, and White, William C.: An Audio Input-Output Computer System for Medical Information. <u>Proceedings of 24th Natiobal Conference, Association for Computing Machinery</u>, August 1969.

Otten, Michael, and Pacak, Milos: Intermediate Languages for Automatic Language Processing. <u>Third International Symposium on Computer and Information Sciences</u>. December 18-20, 1969.

Pacak, Milos, and Pratt, A. W.: The Function of Semantics in Automated Language Processing. <u>Linguistic Applique</u>. (Not published yet)

Pratt, A. W., and Pacak, Milos: Automated Processing of Medical English. <u>International Conference on Computational Linguistics</u>, September 1-4, 1969.

Shapiro, Marvin: Design of an On-line Computer System for Acquisition of Laboratory Data. 12-9-68. <u>Council to Advance Programming</u>, San Francisco.

Sheiner, Lewis B.: Computer-Aided Long-Term Anticoagulation Therapy. Dec. 1969. <u>Computers and Biomedical Research</u>, Volume 2, Number 6.

Simon, Richard: Subroutine Library Organization for Efficient Linking. <u>Journal of the Association for Computing Machinery</u>. (Not published yet)

Simon, Richard, and Lee, Richard C.: The Use of Series-Parallel Graphs in Heuristic Problem Solving. Journal of the Association for Computing Machinery. (Not published yet)

Syed, Daniel, and Kempner, Kenneth M.: Computers and Medical Research. The New Physician. February 1970.

Ulatowska, Hanna K., Scott, Winfield H., Wynn, Lyman C.: Recognizing and Interpreting: A Differentiation of Approaches to Rorchach Responding. Journal of Mental and Nervous Diseases. (Not published yet)

White, William C., Allen, Scott I., Otten, M., Swarthe, Eric: An Experimental Computer Network for Medical Data Processing. Methods of Information in Medicine. July 1969.

Wilson, John, Syed, Daniel, Holsinger, William, DeLeo, James, and Krichevsky, Micah: On-line Data Acquisition for a Bio-Medical Research Laboratory. Council to Advance Programming, Fall of 1968.

Serial No. DCRT 4-1
Computer Systems Laboratory
Systems Design
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer System for NIDR

Previous Serial Number: 4.1

Project Leader: Daniel Syed

Project Description:

Objectives:

The use of the computer as an adjunct to laboratory procedures in the National Institute of Dental Research is intended to both augment and expedite the biochemistry and neurophysiology research programs of that Institute. Specifically, the computer system will modernize techniques for acquiring data from laboratory instruments and will provide for the real-time analysis of data from selected experiments. Ultimately, it is planned to use the computer to control experiments in real-time.

Background:

Subsequent to a detailed requirement study started in February 1967, a process control class computer system was designed and specifications generated. Open competition procurement resulted in the selection of a Honeywell DDP-516 computer which was purchased in June 1968 and installed in the Dental Institute, Building 30, in July 1969.

FY 1970 Activities:

During the past year hardware interfaces and data acquisition programs have been developed for amino acid, scintillation counting and neurophysiological experiments. Some applications programs have been written, however, primary emphasis was devoted toward implementing I/O drivers for non-standard peripherals, establishing a DDP-516 program library on the IBM 360 computer and modifying the real-time monitor of the DDP-516 computer to perform as a stand alone system. The original decision to implement a central facility link via emulation survived an intensive re-evaluation and development of the link is now underway.

Future Efforts:

Although the system should prove to be of some use during the coming year, full effectiveness will not be achievable until additional complements of experiments are implemented and a user oriented operating system is developed.

Serial No. DCRT 4.2
Computer Systems Laboratory
Laboratory Applications
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer System for NIAMD

Previous Serial Numbers: 4.10, 4.21

Project Leader: Marvin Shapiro

Project Description:

Objectives:

The NIAMD computer, located in Building 2, is a real-time data acquisition system. It will enable scientists to improve their research in a number of ways. Data will be collected much more quickly and accurately than was possible before automation and the data will then be available in a form suitable for further processing. This will represent a dramatic reduction in the time required to complete calculations and to plot results, in some cases a reduction from a week to as little as 10 minutes. Also, the scientist should only have to leave the laboratory once, to pick up the results. These improvements, plus the ability to use the computer for feedback control of some experiments, should result in a significant change and improvement in the experimental procedures currently being used by many scientists in NIAMD.

Background:

A computer system for real-time data acquisition was designed in early 1968 and a system meeting the requirements (a Honeywell DDP-516 computer) was ordered in June of 1968. The system was delivered in July 1969, three months later than expected. NIH work on interfacing instruments to the computer and on software modifications to the monitor supplied by Honeywell, began in July 1968.

FY 1970 Activities:

During the past year, phase one of the project was completed. This represented interfacing four instruments (two spectrophotometers, a spectropolarmeter, and a RAMAN spectrometer) to the computer and implementing major software improvements to the manufacturer-supplied software system. These NIH software modifications, necessary to speed up interrupt response and to allow an interactive dialogue between the scientist in the laboratory and the monitor, were completed in March 1970. A large number of programs were written to aid in checkout and in use of the system and special programs for data processing and plotting were written. In addition to hardware

interfacing work, a special Data Acquisition and Display Subsystem was designed and built for conditioning signals entering and leaving the computer and to assist in system checkout.

Future Efforts:

During the coming year work will be done on phase two of the project, which will include the following:

a. Interfacing of more instruments.

b. Connection to two small computers, located in Building 2, to provide them support in terms of added computational power and backup storage.

c. Connection to the central NIH computer system via telephone lines.

d. The development of feedback control of certain applications.

Serial No. DCRT 4-3
Computer Systems Laboratory
Laboratory Applications
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: NHLI Mass Spectrometer Computer System

Previous Serial Numbers: 4.6, 4.11

Project Leader: Eli Gilbert

Project Description:

Objectives:

The National Heart and Lung Institute Mass Spectrometer Computer System located in Dr. Fales' laboratory is designed to collect and process high and low resolution mass spectrometer data. The output from the system is used in the elucidation of the structure of complex molecules. The computer provides a much higher level of performance than was previously obtainable with the mass spectrometer system alone. Complete mass assignments for over 200 masses can now be done, to 6 digit accuracy, in minutes, as compared to many hours of tedious hand work without the aid of the computer. Also, the digitization of the data provided by the computer system enables further processing to be done on the data on a larger computer.

Background:

The MS-9 high resolution and the LKB low resolution mass spectrometers were already available for interfacing to a computer. The initial version of the computer system, consisting of a PDP-8I computer with 4K of memory, a teletype, and A/D conversion equipment, was delivered and interfaced to the MS-9 in December 1968. The incremental plotter was delivered in February 1969 and a high-speed paper tape reader-punch, an incremental tape and a disk were added in June 1969, to complete delivery of the system. The LKB low resolution spectrometer was connected to the computer in February 1970. In March 1970 the system software called SERF, which utilizes the disk, was delivered.

FY 1970 Activities:

The bulk of the work done in the past year has been directed toward making the system operational. Many problems in three main areas--software, electrical noise, and MS-9 magnet instability--have made the system very ineffective. Most of the system noise and spurious interrupts arising from magnet circuit sparking, has been eliminated.

The magnet has been stabilized to the point where it can now produce the desired accuracy and a number of software problems, mainly associated with standardization and automatic starting, have been eliminated.

Future Efforts:

It is hoped that the system will be fully operational in the coming year and that a typical week might see a throughput of as many as 20 samples (as opposed to the past rate of less than one sample) per week.

In addition to providing for routine collection of data, we will be initiating, or continuing, a number of projects. Analysis of peaks from metastable ions will be possible when the analog circuitry (designed and built in CSL) is added to the system and the associated software is working. A mass marker feature being added to the LKB will provide useful plotted output. Some simple feedback control can be provided for spectrometer control with the addition of digital-to-analog converters to the system. Finally, a project may be begun which utilizes mass spectrometer data to identify poisons and barbiturates.

Serial No. DCRT 4-4
Computer Systems Laboratory
Processor Design
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer System for Gerontology Research Center

Previous Serial Number: 4.24

Project Leader: Perry S. Plexico

Project Description:

Objectives:

The recently installed computer system at the Gerontology Research Center, NICHD in Baltimore is intended to accommodate various on-line control and data-acquisition experiments and off-line data processing tasks in support of research by the Center into the nature of the ageing process. The research efforts are multi-disciplinary and encompass the techniques of biochemistry, molecular biology, physiology, and psychology; hence, the applications for which the computer is to be used are expected to be widely varied. Initially the system is to be used for off-line processing, for on-line control of blood glucose and insulin infusion and for on-line psychological and problem solving experiments.

Background:

In February, 1968, the Gerontology Research Center requested that the Computer Systems Laboratory, DCRT provide assistance in the selection of a computer system meeting GRC research requirements. A study of GRC research needs was undertaken and resulted in the preparation of detailed computer system specifications which were used in March, 1969 to solicit proposals from computer manufacturers. On the basis of the recommendation of an evaluation committee composed of representatives from both CSL and GRC, a contract was negotiated in June, 1969 with the Raytheon Company for the procurement of a computer system.

FY 1970 Activities

Primary efforts on this project by CSL staff during the past year have been concentrated in three areas: (a) Study of the computer system's multiprogramming operating system to determine how best to use it in support of GRC research projects, (b) Development of algorithms and programs for an on-line glucose-insulin infusion control experiment, (c) Planning and execution of system test procedures necessary for final acceptance. The system was delivered and installed in early March, 1970. Prior to delivery, efforts had been directed to develop

the software for a glucose/insulin control experiment. In this experiment the computer is to be used in a closed loop control configuration in which it controls the infusion of glucose into the bloodstream of a subject while monitoring blood glucose level. The objective is to discover the control procedures necessary to maintain an above normal glucose level in order to achieve a better understanding of the underlying causes of diabetes.

Future Efforts:

Upon completion of acceptance testing, final checkout of the glucose infusion experiment software will be undertaken. Implementation of on-line experiments in psychological testing and problem solving will proceed concurrently. In addition, a program of indoctrination and training of GRC personnel in the use of the computer system, both for on-line experimentation and off-line data processing, will be instituted.

Serial No. DCRT 4-5
Computer Systems Laboratory
Processor Design
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer Aided Nuclear Magnetic Resonance Spectroscopy

Previous Serial Numbers: None

Project Leader: Perry S. Plexico

Project Description:

Objectives:

The Physical Sciences Laboratory, DCRT, in collaboration with the National Institute of Arthritis and Metabolic Diseases, plans to use a pulsed nuclear magnetic resonance (NMR) spectrometer in conjunction with a small-scale on-line computer system for the performance of high resolution Fourier transform spectroscopy. NMR techniques are particularly valuable in the field of biochemistry for studying the fundamental properties of structure, binding and conformation in biomolecules and biopolymers. The advantage of Fourier transform spectroscopy, as compared to conventional frequency sweep techniques, is that significant improvement in signal-to-noise ratio can be realized in appreciably less time. This is an important factor, both in terms of manpower expenditure and in circumventing the problem of inherent instrument instability.

Background:

In August of 1969, the Computer Systems Laboratory, DCRT, made a study of the computer requirements for NMR Fourier transform spectroscopy in which both the possibility of using the existing NIAMD Honeywell 516 computer and of procuring a new dedicated system were considered. It was determined that because of the high data rates and memory requirements of the NMR experiments, scheduling difficulties with existing facilities would be extremely detrimental to other projects. Furthermore, a number of modifications, at substantial cost, would have been necessary for the Honeywell 516 configuration; hence, the dedicated system approach was recommended.

FY 1970 Activities:

Development of detailed specifications for this system began in November 1969. Upon completion of these system specifications and receipt of departmental clearance, proposals were solicited from manufacturers in April 1970, for a computer system for NMR Fourier transform spectroscopy. Since the computer is to be located at a site some distance from the spectrometer, work has also begun on the

development of data transmission techniques to accommodate communication between the two machines.

Future Efforts:

Following receipt of proposals from manufacturers, an evaluation committee comprised of members of both the Computer Systems Laboratory and the Physical Sciences Laboratory will select a system from those proposed. Once the computer configuration is known, development of software for data acquisition and Fourier transformation will be undertaken. It is hoped that completion of this work will coincide with the estimated system delivery date of December 1970, so that NMR experimental work can begin soon thereafter.

Serial No. DCRT 4-6
Computer Systems Laboratory
Systems Design
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer System for NIMH

Previous Serial Number: 4.23

Project Leader: Victor Colburn

Project Description:

Objectives:

This project will provide a real-time computer system for the support of research studies conducted by scientists of the National Institute of Mental Health located in Building 10. The system will provide real-time data acquisition, analysis and control for experiments in Problem-Solving, Learning, Evoked Response, Perception, Sleep and Dreaming, Patient/Family Interaction and others. It will also support experiments which relate analysis of electroencephalographs and other physiological variables to clinical diagnosis and classification of NIMH patients. When time permits, the system will also be used for certain off-line support of these same scientific activities.

Background:

This project originated with an initial request for assistance from NIMH and subsequent study of their requirements by CSL during calendar 1967. Approval for computer procurement was received from HEW in February 1969. The procurement was advertised March 12, 1969; and evaluation of proposals was completed during May 1969. A fund squeeze frustrated straightforward procurement at this point, nevertheless, a contract was finally signed on November 21, 1969 with Systems Engineering Laboratories of Fort Lauderdale, Florida.

FY 1970 Activities:

With the unexpected cutback of funds on this project, it became necessary to tailor the NIMH requirements to a two phase procurement to better match the available funds. The first phase was to be supported in FY-70 with about $100,000. Careful re-evaluations of both the NIMH requirements and the contractor proposals, were conducted prior to final negotiations. The contract signed with S.E.L. of Florida on November 21, 1969 was for a dollar amount of $100,309 and provides for a first phase computer system of about 70 percent of the dollar value of the intended total system. About four tenths of this procurement went for the purchase of equipment--the rest of the money for the lease of equipment. Delivery is scheduled for spring 1970.

NIMH has assumed responsibility for all hardware aspects of the system and for its operation and maintenance. NIMH will further provide essentially all scientific user programs. DCRT has assumed system implementation responsibility which involves:

1. extending the real-time Monitor (system software) from its initial state of single-user ("multiprogramming") system to a multi-user system which is both flexible and secure, and

2. achieving satisfactory overall system performance.

Future Efforts:

CSL effort over the coming year will be directed to the following:

1. Provide for initial operation of the system using the existing system software.

2. Implement the system software extension which will provide the remote user with control of the system.

Serial No. DCRT 4-7
Computer Systems Laboratory
Medical Information Science
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Medical Telecommunications

Previous Serial Numbers: 4.3, 4.4, 4.7, 4.18, 4.20

Project Leader: William White

Project Description:

Objectives:

The objectives of this project are: 1) to facilitate use of the computer for the practicing clinician, medical research physician and supporting paramedical personnel by developing convenient, economical and reliable methods for remote access to computer-stored information and computations, and 2) to develop useful computer programs to assist the health professional in patient-care decision-making.

Background:

This project began in 1968 as an outgrowth of a project to develop a CRT terminal oriented clinical information system for the Clinical Center. We realized that available terminals were not optimum for the most common clinical uses, and that economic constraints required a different approach than the one we originally proposed. We decided to explore use of the telephone as a computer terminal and commercial time-sharing computer services for use by the practicing physician. Unfortunately, use of available commercial systems is complicated by the fact that each has its own idiosyncracies of input and output, thus making use of it by the non-computer oriented physician unattractive. In addition, no single commercial time-sharing system is adequately reliable. We hoped these two problems could be overcome by using an additional small computer to translate each commercial computer's format to a simple format familiar to the physician, and to allow for automatically switching from one commercial system to another so that even when any given commercial system was inoperative, reliable service would be available to the physician user. In order to establish the feasibility of this approach and to establish the utility of terminals such as a Xerox telecopier and the touch tone telephone in this health care oriented application, use was made of a HP2116B computer which was temporarily available to us. This trial proved so promising that we competitively procured an SEL 810B computer in the spring of 1969 to further explore these ideas.

FY 1970 Activities:

During FY 1970 a number of medically oriented programs were written for medical telecommunications application. These programs include a pediatric burn therapy computation, an intravenous drug compatibility table lookup, a pill and poison identification program, a digitalis dose calculation program, an antibiotic dosage calculation program and some diagnostic assistance programs. The SEL 810B computer was installed and development of the systems software to accommodate touch-tone telephone input and voice response was started. The system was interfaced with the DCRT PDP-10 to provide for computational support, auxiliary storage and the use of the PDP-10's peripheral equipment. Design and construction of a voice response unit to accommodate four input/output lines using delta-modulation data compression was completed.

Future Efforts:

More development is required before the system can be called operational. During the coming year we expect to continue development of this system by: further software development, writing more applications programs, soliciting and translating for use on this system existing medical programs developed at other institutions; and by further developing terminals more suited for use by the physician including those which can accommodate hard copy.

Serial No. DCRT - 4-8
Computer Systems Laboratory
Systems Design
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Cardiovascular Studies

Previous Serial Numbers: 4.9, 4.15

Project Leader: Daniel Syed

Project Description:

Objectives:

The application of computers to the analysis of the complex signals obtained from the cardiovascular system can significantly enhance the benefits derived from basic clinical research, aid in the diagnosis of cardiovascular deficiencies, and contribute to patient care through continuous monitoring of cardiovascular events.

Background:

Early efforts centered around the development of hardware and software techniques for the detection and analysis of cardiac arrhythmias utilizing a single electrocardiographic signal. Techniques for the transmission of ECG's throughout the NIH and subsequent digitization at the site of analysis were developed during this period. In addition, the analysis of aortic pressure was also initiated. The general goal of an integrated cardiovascular analysis system was proposed, with emphasis upon the eventual capability for hardware preprocessing of the ECG, aortic pressure waveform, and the phonocardiogram signals.

FY 1970 Activity:

During FY 1970 further improvements were added to the hybrid computer arrhythmia monitoring system, and an extremely reliable hardware ECG preprocessor was completed. An identical version of the hardware ECG preprocessor was also added to the original ECG arrhythmia analyzer built in FY 1969 and reported on in that year. A hybrid computer software package was developed to allow the evaluation of the pulse contour method of stroke volume determination. Simultaneously, a hardware aortic pressure preprocessor was designed and constructed, and development was initiated on a small battery powered acoustic coupler to allow simplified ECG transmission from fingertips to any remote site, via a standard telephone.

Additional software systems were developed to support several research programs underway within the Surgery Branch of NHLI. These programs allow the evaluation of cardiac contractility as reflected in such measurements as rate of ventricular pressure rise (with Dr. Reis) and motion of the heart border (with Dr. Levitsky).

Future Efforts:

Further development is required to optimize the hardware systems for arrhythmia analysis and aortic pressure preprocessing. Construction of a phonocardiogram preprocessor is to be started in the near future. These devices will be proved in a thorough clinical evaluation program. The post-operative recovery room of the NHLI Surgery Branch will be equipped with data phone lines to allow utilization of the cardiac contractility programs in an on-line environment during forthcoming clinical studies.

Serial No. DCRT 4-9
Computer Systems Laboratory
Medical Information Science
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Automated Processing of Medical English

Previous Serial Number: 4.19

Project Leader: Milos Pacak

Project Description:

Objectives:
The major objectives of the project are to develop the methodology for machine encoding of diagnostic statements into a file, and to provide the capability to retrieve relevant information from the data file with a high degree of completeness.

Background:
The program for automated processing of medical English was started about two years ago. The first phase of the project was the development of a program by which medical diagnosis will be indexed by the computer and stored for retrieval of relevant information. The retrieval program developed by M. Epstein is already operational. For details see the manual "An Information Processing System for Pathology Data", NIH, DCRT.

FY 1970 Activities:
During the past year the main effort was concentrated on improving components of the system for automated indexing and editing SNOP to bring it into conformance with empirical results in close cooperation with the Department of Pathology. It is expected that the automated indexing program will be tested on a large scale in the immediate future. In the initial stage some postediting of the input data will be required before the system becomes fully automatized.

A theoretical basis was laid for the development of a program which is intended to handle general medical English. The emphasis was put on the design of a semantic model for automated processing of medical English. For more detailed explanation see the paper: "Automated Processing of Medical English".

Research is being carried out to develop a procedure for morphosemantic analysis of compound medical terms which are derived from Greek and Latin. The procedure, when it is developed, should serve as a basis for automatic construction of medical micro-glossaries.

In addition, a critical comparison of available authority files for the control of medical report literature is underway.

It is expected that further experimentation with the system will eliminate the need for hand coding of diagnoses and, at the same time, provide us with more data for additional improvements.

Work to date has proceeded in an exploratory mode. The primary emphasis has been on the development of a linguistic model. Programs were written to test components of the model. We anticipate that experience in an operational setting and closer exchange of ideas with the user at this level will suggest improvements for overall efficiency.

Future Efforts:

The experimental design of a formal semantic model for medical English is going to continue with the aim to be implemented in a broader medical context. One of the main objectives of the project will be the preparation of an experimental program for automated build-up of computer-oriented medical micro-glossaries, the lack of which is felt by the medical community in the U.S. and abroad.

Serial No. DCRT 4-10
Computer Systems Laboratory
NA
Bethesda

PHS-NIH
Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer Systems Laboratory Consultation

Previous Serial Number: 4.12

Project Leader: Alan Demmerle

Project Description:

Objectives:
This project incorporates a number of consultation activities directed toward providing requesting Institutes with guidance in any aspect of computer systems design and use.

Background:
The project seeks to afford assistance to the intramural and extramural programs of all Institutes, however, to date, consultative activities have been directed primarily toward the support of the data management portions of large dollar value, long term contracts and grants sponsored by the NHLI. During collaboration with NHLI, efforts have included providing contractor guidance, designing a complete system or simply offering advice to NHLI contract office.

FY 1970 Activities:
During the past year effort was concentrated on the support of two major NHLI projects, the Myocardial Infarction Program and the Blood Resources Program. Activities in support of the former program culminated in a requirement study, computer system design and procurement evaluation. An important adjunct to this MIRU effort was continued contractor guidance and project coordination of five data management centers. The primary contribution in the Blood Resources area has been to evaluate contractor performance and to help shape future extensions of current contracts.

Future Activities:
Future efforts in this project are determined by ad hoc requests from the Institutes.

July 1, 1969 through June 30, 1970

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH

DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT - 5

2. PHYSICAL SCIENCES LABORATORY — 3. Dr. G. H. Weiss

I. SUMMARY

The Physical Sciences Laboratory is devoted to the study of problems in physics and chemistry that relate to the biological sciences. Several disciplines are represented in the membership of the laboratory. These include applied mathematics, theoretical chemistry, and theoretical physics. Whenever possible the theoretical studies are performed in conjunction with experimental work, either in collaboration with workers in outside units, or by members of the Physical Sciences Laboratory working in other laboratories at NIH. In addition to performing reseach of its own choosing, members of the Physical Sciences Laboratory provide consultation services to other researchers at NIH on different topics in the disciplines represented in the Laboratory. These will be enumerated in more detail below.

II. CURRENT BRANCH PROGRAMS

A. Objectives

Members of the Physical Sciences Laboratory are presently pursuing many lines of investigation into problems of theoretical physics, theoretical chemistry, and the application of numerical techniques to the processing of data from biochemical experiments. During the past year the Laboratory has included between 9 and 11 professional workers. The most notable achievement by members of the Laboratory this past year has been the application of a sophisticated statistical mechanical theory of the forces between close objects (known technically as Van der Waals forces) to cellular phenomena. The theory developed in the course of the investigation has led to a qualitative and partly quantitative understanding of such phenomena as cell aggregation and contact inhibition, and promises to provide a unified description of many previously isolated biological facts.

B. Progress of Current Programs

1. Development of Ultracentrifugation Theory

a. An approximate theory of boundary spreading in the ultracentrifuge has been developed using the methods of singular perturbation theory. These results have been compared to accurate solutions generated numerically for sedimentation with pressure dependence, with good agreement.

b. A new theory has been developed for predicting the effects of pressure dependence on velocity sedimentation, with better agreement with qualitative effects than the presently accepted theory.

c. A method has been devised for measuring the diffusion coefficient of small molecules during the acceleration phase of the experiment. The theoretical work has been confirmed by a concurrent experiment.

2. Theory of the Helix-Coil Transition of Polypeptides

a. The results of NMR experiments on polypeptides undergoing a transition have been in excellent agreement with the theory developed earlier. In particular the experiments confirm predictions of different long- and short-chain length dependence in NMR spectra over the transformation region.

3. Computer Analysis of NMR Spectra

The program to analyze NMR spectra in terms of the chemical shift and spin-spin interaction parameters has been completed, debugged, and applied to experimental data.

4. Molecular Model Building

Computer programs have been written to display molecular structure on a CRT for the AGT-30.

5. Biophysical Analysis

Results have been obtained for the equilibrium solutions, and for the rate of approach to equilibrium, predicted by the kinetic theory of the helix-coil transition of small macromolecules with limiting nucleation steps.

6. Configurational Statistics of Polymer Chains

The theory of the excluded volume effect in macromolecules is being extended to determine the dependence on size of the mean square end-to-end distance. The program has been set up in outline and some of the heavy calculations performed.

7. Excitation and Transport Properties of Fluids

Instrumentation of a laser light scattering spectrophotometer is being developed under contract with MIT. Preliminary experiments indicate the feasibility of using this apparatus for measuring the diffusion coefficients of biologically interesting molecules.

8. Intermolecular Forces in Biological Structures

The method of Lifshitz has been applied to the calculation of Van der Waals forces in biologically interesting structures. It has been shown that many biological phenomena on the cellular level can be explained in a

relatively simple theoretical way by the results implicit in this theory. A very important finding is that the "fuzzy" layer around the cell surface is a very important determinant of the interactive properties of cells.

9. Consulting Services

Curve fitting techniques have been used to analyze transient experiments in enzyme kinetics, and to determine rate constants.

10. Fundamental Studies

Research in the theory of clinical trials is being pursued. Results have been obtained in the comparison of vector-at-a-time sampling and play-the-winner sampling for choosing the better of two binomial populations. The results indicate that neither method is uniformly best for all values of success probability.

11. Computer-Assisted NMR

Curve fitting techniques developed in the Physical Sciences Laboratory have been successfully applied to the decomposition of NMR spectra. Relative peak areas have been determined quite accurately and the chemical shift of the imidazole resonances as a function of pH have been fitted to their theoretical values.

12. High Resolution Carbon-13 NMR Spectroscopy at 55 MHz

Preliminary experiments indicate that a 100-fold enhancement in the carbon-13 signal/noise ratio is possible over conventional techniques, for a given observation time.

13. Rate Studies by NMR Spectroscopy

Accurate thermodynamic information on activation energies can be obtained by the use of NMR spectroscopy and the use of numerical curve fitting techniques.

14. Applications of Walsh Transforms in NMR Spectroscopy

The project has just begun, and no results are available so far.

15. Models for the Transport Properties of Membranes

Results have been obtained relating the (inverse power) force laws between ions and membranes and the transport properties of the membranes.

Serial No. 5.1
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 20, 1970

Project Title: Theory of the Ultracentrifuge

Previous Serial Number: 5.1

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: Ralph Nossal, Ph.D.

Cooperating Units: Werner Klee, Ph.D.,NIMH

Man Years

Total:	.60
Professional:	.50
Other:	.10

Project Description:

Objectives:

To determine the effects of various factors such as concentration dependent sedimentation, pressure, density gradients, variations in rotor speed, and polydispersity on current techniques for determining molecular weights. To devise corrections and new techniques of ultracentrifugation which bypass or eliminate these effects.

Methods:

The methods employed involve analytical and numerical solutions of linear and nonlinear partial differential equations.

Major Findings:

A singular perturbation method for the solution of linear parabolic partial differential equations has been devised. This method has been applied to the solution of the Lamm equation for ultracentrifugation, taking pressure effects into account. The agreement with accurate numerical solutions is extremely good, so that one now can calculate boundary spreading effects due to diffusion. A technique has been devised for

measuring the diffusion constant for small molecules during the acceleration period. The method is much quicker than any other method using the ultracentrifuge.

Significance to Biomedical Research

Most techniques for reducing ultracentrifuge data from velocity experiments rely on a theory that neglects diffusion. The approximate theory of boundary spreading will allow an estimate of how reliable such theories are, and when the results of such theories can be used. The measurement of diffusion constants allows a rapid, easily implementable method for the determination of these constants for small molecules, for which earlier ultracentrifugal methods were not useful. Since the ultracentrifuge is an instrument found in every biochemistry laboratory, the development of methods for data reduction is of some importance in the analysis of many molecules of biological interest.

Honors and Awards: None

Publications:

Nossal, R. J. and Weiss, G. H.: Early time expansions for sedimentation of weakly non-ideal solutions. Biopolymers 7, 353-360, 1969.

Weiss, G. H., Dishon, M., Yphantis, D. A.: Numerical solutions of the Lamm equation. V. Band centrifugations. Annals of the New York Academy of Science, 164, no. 1, 33-51, 1969.

Weiss, G. H. and Dishon, M.: Diffusional boundary spreading in the ultracentrifuge. Biopolymers (to appear).

Serial No. 5.2
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Theory of the Helix-Random Coil Transformation of Polypeptides in Solution

Previous Serial Number: 5.2

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: B. W. Ninham, Ph.D., V. A. Parsegian, Ph.D., and J. Milstein (NIAMD)

Cooperating Units: Laboratory of Physical Biology, NIAMD

Man Years:

Total:	.50
Professional	.50
Other	.00

Project Description:

Objectives:

To gain insight into the mechanisms of the helix-random coil transformation, protein denaturation, and enzyme activity and to formulate a model of the transformation based on both NMR spectroscopic and rapid reaction kinetic techniques.

Methods:

1. Application of NMR spectroscopy as the experimental approach to obtain lifetime information associated with the helix and random coil.

2. Examination of an exactly solvable kinetic model of the transformation where both nucleation and chain length effects are included.

Major Findings:

High-resolution NMR spectra at 220 MHz obtained on samples of poly-γ-benzyl-L-glutamate of different molecular weights show separate

helix and random coil peaks and also that the transformation is strongly dependent on chain length. Furthermore, the kinetic model of the process agrees with the kinetic and NMR times and also correctly predicts qualitatively different long and short chain length behavior in the NMR spectra over the transformation region.

Publications:

Ferretti, J. A. and Ninham, B. W.: Nuclear magnetic resonance investigation of the helix to random coil transformation in poly-(α-amino acids). II. Poly(γ-benzyl-L-glutamate). Macromolecules 3, 30-33, 1970.

Ferretti, J. A., Ninham, B. W. and Parsegian, V. A.: Theory of the helix-random coil transformation. Macromolecules 3, 34-42, 1970.

Serial No. 5.3
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer Analysis of Nuclear Magnetic Resonance Spectra

Previous Serial Number: 5.3

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Rolf B. Johannesen (NBS)

Man Years:

Total:	.30
Professional	.30
Other	.00

Project Description:

Objectives:

To analyze nuclear magnetic resonance spectra in terms of the chemical shift and spin-spin interaction parameters.

Methods:

An iterative computer program was written to analyze complex NMR spectra.

Major Findings:

It is possible to analyze very complex molecules containing up to between twenty and thirty spins quite efficiently.

Significance to Biomedical Research:

Nuclear magnetic resonance techniques have recently been bery effectively applied to biologically interesting molecules. Since these molecules tend to be fairly complicated, sophisticated methods of data reduction are required.

Honors and Awards: None

Publications:

Ferretti, J. A., Johannesen, R. B., and Harris, R. K.: VEAITR: A new computer program for analysis of the proton spectrum of triisopropylphosphine. Journal of Magnetic Resonance, (to appear).

Serial No. 5.4
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Molecular Mechanics

Previous Serial Number: 5.4

Principal Investigator: William P. Minicozzi, M.D.

Other Investigators: Michael Stroot

Cooperating Units: None

Man Years:

Total:	1.0
Professional:	1.0
Other:	0.0

Project Description:

Objectives:

To derive molecular interaction energy functions from experimentally determined observables (obtained from molecular systems in the gas and crystal phases). These functions can be used to predict pertinent physical quantities of biologically important systems. Our ultimate goal is to make feasible the understanding and solution of biological problems on a detailed molecular basis.

Methods:

Using physical principles, equations to calculate quantum mechanical observables are derived. Mathematical methods to evaluate these equations in a feasible manner have been developed (where necessary). Programs to perform the above calculations for a general molecule and/or molecular system in the gas or crystal phase have been written. In addition programs to display molecular structures on a CRT have been written for the AGT-30.

Major Findings:

A relatively simple mathematical model, which lends itself to rapid evaluation, can be used to accurately calculate molecular observables in both the gas and crystal phases of a molecular system.

Significance to Biomedical Research:

It is now theoretically and technically possible to determine pertinent molecular characteristics of biologically important molecules in feasible amounts of computer time. This fact makes it possible to investigate, determine, control, and predict the biological activities of molecular compounds as they are determined by molecular interactions. The significance to biomedical research is obvious.

Honors and Awards: None

Publications:

Minicozzi, W. P. and Stroot, M. T.: On the determination of interaction energy functions II. Crystalline formic acid. J. of Comp. Physics, (in press).

Serial No. 5.5
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Biophysical Analysis

Previous Serial Number: 5.5

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: Barry W. Ninham, Ph.D., Michael Mackey, Ph.D. and Robert Zwanzig, Ph.D.

Cooperating Units: Harold Lecar, Ph.D., Gerald Ehrenstein, Ph.D.
Biophysics Laboratory, NINDB

Man Years:

Total:	1.45
Professional:	1.30
Other:	.15

Objectives:

To supply theoretical foundations and experimental modesl for various observations which arise in physiology and biophysical chemistry. The following problems have received particular attention during the past year.

1) Models for macromolecular rate processes: The objective is to provide mathematical foundations for various kinetic phenomena arising in studies of macromolecular biochemistry.

2) Phenomena occuring in phospholipid bilayer membranes: The objective is to develop experiments and theories relating conductivity measurements and transport data to the molecular processes which give rise to excitability in biomembranes.

3) Threshold fluctuations in nerves: The objectives are to relate the fluctuations in firing thresholds of nerves to the chemical and physical processes underlying excitation and to analyze existing data in order to test theories concerning mechanisms of transport of ions across nerve membranes.

Methods:

1) Kinetic equations have been derived and investigated for a number

of possible association mechanisms.

2) Laboratory technique of microbiology and electrophysiology are combined with techniques of mathematical physics.

3) Existing neurophysiological equations (the Hodgkin-Huxley equations) have been modified to include fluctuating forces. Methods of non-linear mechanics and statistical physics have been applied to analyze the equations and digital computers are used to determine phase trajectories.

Major Findings:

1) Results have been obtained for the equilibrium solutions and approach to equilibrium of kinetic equations describing

a) the helix-coil transition of finite length macromolecules with limiting nucleation processes, and b) rate limited surface adsorption processes.

2) The properties of individual voltage dependent transport units derived from aerobacter cloacae have been determined and related to the properties of membranes having many such units.

Significance to Biomedical Research:

1) Kinetic models: The theoretical studies describe systems which are analogs of various physiological phenomena.

2) Reconstituted phospholipid membranes: The membrane components can be controlled by the experimentalist, thus providing an opportunity to devise procedures and test theories applicable to natural biological membranes.

3) Threshold fluctuations: It is believed that the study will facilitate better understanding of the physical processes underlying the excitation and propagation of nervous impulse.

Publications:

Ninham, B. W., Nossal, R. and Zwanzig, R.: Kinetics of a sequence of first order reactions. J. Chem. Phys. 51, 5028-5033, 1970.

Nossal, R. and Ninham, B. W.: Kinetic equations for surface adsorption. Biopolymers, 9, 103-111, 1970.

Ehrenstein, G., Lecar, H. and Nossal, R.: The nature of the negative resistance in biomolecular lipid membranes containing EIM. J. Gen Physiol. 55, 119-133, 1970.

Serial No. 5.6
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report

Project Title: Configurational Statistics of Polymer Chains

Previous Serial Number: None

Principal Investigator: Ivan Darvey, Ph.D.

Other Investigators: Barry W. Ninham, Ph.D.

Man Years:

Total:	.20
Professional:	.20
Other:	.00

Objectives:

To calculate the mean squared end to end distance of a polymer chain taking into account the excluded volume effect.

Methods:

Diagrammatic methods as used both in quantum field theory and other many body problems.

Major Findings:

Only preliminary calculations of the fundamental series of diagrams have been made. Detailed study of the asymptotic properties of the resulting integrals is in its early stages.

Honors and Awards: None

Publications: None

Serial No. 5.7
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Reports
July 1, 1969 through June 30, 1970

Project Title: Excitation and Transport Properties of Fluids

Previous Serial Number: 5.7

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years:

Total:	.20
Professional	.20
Other:	.00

Project Description:

Objectives:

To provide basic knowledge concerning the excitation properties and transport properties of both simple fluids and complex solutions containing biological macromolecules.

To develop a laser light scattering spectrometer to measure hydrodynamic coefficients of biological macromolecules and the rate constants of biomolecular reactions.

Methods:

Theoretical techniques of mathematical physics and statistical mechanics are employed in order to develop new physical theories. Theoretical studies are performed in support of new experiments, particularly those involving the scattering of light from laser sources. A correlation function laser scattering spectrometer is being developed under contract.

Major Findings:

Preliminary data prove the feasibility of using the laser light scattering correlation spectrometer for the purpose of measuring

the diffusion coefficients of biological molecules having a wide range of sizes (even as small as MW 1000). Additional data acquisition channels are currently being constructed for the spectrometer.

Significance to Biomedical Research:

Almost all biological phenomena occur in a fluid environment. A number of fundamental questions concerning the physical behavior of fluids yet remain unanswered. Their elucidation will ultimately facilitate better understanding of the functions and properties of biological systems. The development of the laser correlation spectrometer may enable rapid and precise measurement of various physical parameters characterizing systems of biological molecules.

Honors and Awards: None

Publications: None

Serial No. 5.8
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Intermolecular Forces in Biological Structures

Previous Serial Number: 5.8

Principal Investigators: V. Adrian Parsegian, Ph.D. and B. W. Ninham, Ph.D.

Other Investigators: None

Cooperating Units: R. Woodman, Microbiological Associates

Man Years:

Total:	1.5
Professional:	1.5
Other:	0

Project Description:

Objectives:

To identify and learn to calculate those intermolecular forces governing the structure of matter at the biological level. These have been Coulombic (electrostatic) interactions between charged species and Van der Waals (electrodynamic) forces.

Methods:

Classical and quantum-mechanical treatment of electromagnetic and statistical-mechanical behavior as well as molecular models of specific structural interactions.

Major Findings:

We have successfully applied and generalized the very powerful method of Lifshitz to calculate the Van der Waals force of attraction in the following situations:

(a) Water across a thin lipid film.
(b) Pressure across a triple layer soap film.
(c) Cell-cell attraction forces at close contact.

(d) multilayer systems specifically myelin figures.
(e) Protein-protein molecule interactions in dilute solution.

Significance to Biomedical Research

This is the first application of this or any other rigorous method to problems of biological importance. The theory correctly predicts the experimentally observed strength of dispersion force. The method has opened up a new way of thinking about biological structure. Several new ideas have emerged from our analysis.

1. The cell-cell attraction in tissues is probably dominated and determined by the saccharide-protein "fuzz" coating the cell. Cell contact specificity may be related to fuzz optical properties.

2. The absorption spectrum from all frequencies -- microwave through ultraviolet -- is important in determining Van der Waals forces (previously all but the ultraviolet was neglected).

3. "Hydrophobic" bonding is probably a low frequency Van der Waals force.

4. Van der Waals forces are explicitly temperature-dependent in biological systems.

5. "Kirkwood-Shumaker" proton fluctuation forces between proteins in solution are as important as ultraviolet frequency Van der Waals forces.

We now have a systematic means of designing drugs to control cell-cell contact and are applying this to the experimental aggregation of L1210 cells.

The activity of cell surface enzymes is controlled by the enzyme's ionic microenvironment. Changes in ion concentration upon cell contact can be related to the changes in enzyme activity and hence "contact inhibition" phenomena.

Honors and Awards: None

Publications:

Parsegian, V. A. and Ninham, B. W.: Application of the Lifshitz theory to the calculation of Van der Waals forces across thin lipid films. Nature 224, 1197-1199, 1969.

Ninham, B. W. and Parsegian, V. A.: Van der Waals forces across triple-layer films. J. Chem. Phys. (to appear).

Ninham, B. W. and Parsegian, V. A.: Van der Waals forces: Special characteristics in lipid-water systems and a general method of calculation based on the Lifshitz theory. Biophys. J. (to appear).

Parsegian, V. A. and Ninham, B. W.: Temperature-dependent Van der Waals forces. Biophys. J. (to appear).

Parsegian, V. A.: Dielectric aspects of biological materials. Annual Digest of Literature in Dielectrics, 32, 16-32, 1969.

Serial No. 5.9
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Consulting Services

Previous Serial Number: 5.9

Principal Investigators: George H. Weiss, Ph.D., Ivan G. Darvey, Ph.D., Mildred L. McNeel, Richard I. Shrager.

Other Investigators: Leonard Kohn, M.D., Alan Schechter, Ph.D., Jack Folk, Ph. D., Ronald Chung, Ph.D., Stanley Shackney, M.D., Yashar Hirshaut, M.D., E. Anthony Jones, M.D., Nathaniel Berlin, M.D. and Ayub Ommaya, M.D.

Cooperating Units: Laboratory of Physical Biology, NIAMD; Laboratory of Molecular Biology, NIAMD, Office of the Director, NCI; Sloan-Kettering Institute.

Man Years:

Total:	2.0
Professional:	2.0
Other:	0.0

Project Description:

Objective:

To provide consulting services in applied mathematics, biometry, theoretical chemistry, theroetical physics, and various aspects of numerical analysis to workers primarily in experimental fields.

Methods:

The methods used are mathematical as applied specifically to problems in the physical sciences. Computer programs have been written for nonlinear curve fitting that have been applied to a variety of problems.

Major Findings:

Curve fitting techniques have been used, for the first time, to determine rate constants from the transient phase of kinetic experiments.

Data relating the force of an impinging blow to the probability of concussion has been analyzed, showing a clear exponential dependence.

Experiments at the Sloan-Kettering Institute are being carried out to develop a tissue culture system as a method of bioassay of anti-cancer agents. Results obtained so far indicate that the bioassay is accurate to within 0.4 log dilution with a small number of replications.

Significance to Biomedical Research:

All of the collaborative projects mentioned have been on projects initiated by workers in other Institutes. The output of major significance from this project is in the nature of methodology developed for specific medical or biochemical projects.

Honors and Awards: None

Publications:

Weiss, G. H., Hirshaut, Y., and Perry, S.: The use of long-term leukocyte cell cultures as models for the study of antileukemic agents. Cancer Research, 29, 1732-1740, 1969.

Weiss, G. H. and Fischmann, E. J.: Effect of reducing the number of surface potential measurements on estimates of heart dipole moment. Transactions of the IEEE on Biomedical Engineering, BME-17, 58-65, 1970.

Serial No. 5.10
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Fundamental Studies

Previous Serial Number: 5.10

Principal Investigators: George H. Weiss, Ph.D., Ralph J. Nossal, Ph.D., Barry W. Ninham, Ph.D. and Andrew G. De Rocco, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Hours:

Total:	2.5
Professional:	2.0
Other:	.5

Project Description:

Objective:

This project presently encompasses several lines of investigation, not all of which relate to biomedical problems. These are:

1. The study of statistical procedures for the design and evaluation of clinical trials.

2. A study of perturbation procedures in the mathematical theory of epidemics.

3. A study of mathematical modesl of vehicular traffic flow.

4. The study of molecular forces between pairs of polyatomic systems, with emphasis on the additivity or nonadditivity of such forces.

Methods:

A variety of methods are necessary for the problems mentioned above. These include the theory of stochastic processes, classical analysis, and methods of mathematical physics.

Major Findings:

On the first of the projects mentioned above, attention has been focussed on techniques for choosing the better of two drugs through a clinical trial designed so that the smallest number of patients are tested on the poorer drug, consistent with fixed levels of discrimination and accuracy. Two techniques for sampling have been examined, but neither is uniformly better than the other. On the second we have devised a perturbation method for large populations, that elucidates the transition between the stochastic and deterministic theories, and allows the calculation of correction terms. On the third project we have calculated the effects of different lengths of acceleration lane on merging delays at intersections. On the fourth project we have shown that there appears to be no net contribution to nonadditive forces due to internal coupling for many cluster configurations.

Significance to Biomedical Research:

The work on clinical trials is obviously and directly related to many types of comparative clinical trials. The perturbation method will be useful both in the theory of epidemics and in the treatment of problems in theoretical ecology.

Honors and Awards: None

Publications:

Weiss, G. H.: An introduction to the statistical theory of irreversible processes. In Henley, S. M. (ed.) Transport Phenomena in Fluids. New York, Dekker Publishers, 1969, pp. 75-118.

Weiss, G. H., Shuler, K., and Oppenheim, I.: On the stochastic and deterministic formulation of chemical rate equations. Journal of Chemical Physics 50, 406-413, 1969.

Weiss, G. H. and Zajicek, G.: The kinetics of red blood cells following hemolysis. Journal of Theoretical Biology 23, 475-491, 1969.

Weiss, G. H., Oppenheim, I., and Shuler, K.: On the decay of correlations. II. Relaxation of momentum correlations and momentum distributions in harmonic oscillator chains. Journal of Chemical Physics 50, 2662-2670.

Weiss, G. H.: The intersection delay problem with mixed cars and trucks. Transportation Research 3, 195-199, 1969.

Weiss, G. H.: Contributions to the stochastic theory of chromatographic kinetics. Separation Science 5, 51-62, 1970.

Weiss, G. H. and Gillis, J.: On the expected number of distinct sites visited by a random walk with an infinite variance. Journal of Mathematical Physics (to appear).

Weiss, G. H. and Sobel, M.: Play-the-winner sampling for selecting the better of two binomial populations. Biometrika (to appear).

Weiss, G. H. and Blumenfeld, D.: Routing in a circular city with two ring roads. Transportation Research (to appear).

De Rocco, A. G. and Yorke, E. D.: Non-additivity in polyatomic systems. Journal of Chemical Physics (to appear).

Parsegian, A.: A hybrid bus for intracity transportation. Transportation Research, 3, 307-315, 1969.

Serial No. 5.11
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Computer-Assisted Nuclear Magnetic Resonance (NMR) Studies of Proteins

Principal Investigator: Jack S. Cohen, Ph.D.

Other Investigators: A. Schechter, M.D., M. McNeel and R. I. Shrager

Cooperating Units: None

Man Years:

Total:	.75
Professional:	.75
Other:	.00

Project Description:

Objectives:

1. To gain information on the structure and function of enzymes.

2. To extend the applicability of high resolution NMR to proteins.

Methods:

The region of the NMR spectrum containing the histidine imidazole C2-H resonances of a number of proteins has been subjected to detailed analysis. The 220 MHz time-averaged spectra were digitized and fitted with a series of Lorentzian curves using the MODELAIDE least-squares fitting program with the on-line IBM 2250 display unit. The chemical shifts of the imidazole resonances as a function of pH have been fitted to their theoretical values.

Major Findings:

1. The relative areas of peaks in the spectra can be measured, thus aiding in the identification of imidazole resonances.

2. The data for the four histidines of Staphylococcal nuclease indicate single titration curves.

3. The fitting of the titration data enables the precise determination of the individual histidine microscopic dissociation constants (pK) and allows an objective criterion for deciding the true continuity when curves cross.

4. The findings for S. nuclease do not indicate a conformational equilibrium involving a histidine residue, contrary to the recent interpretation by Markley and Jardetzky of their (100 MHz) NMR data.

Significance to Biomedical Research

1. A general and objective method for the analysis of resolved resonances in protein NMR spectra has been developed. This should lead to more and improved applications of this technique, resulting hopefully in important conclusions regarding protein structure and function.

2. An analysis of the titration data for pancreatic ribonuclease is under way. This approach provides in principle a means of distinguishing between published mechanisms of its activity.

3. To extend the technique to proteins of higher molecular weights, bovine carbonic anhydrase (MW 30,000) is being studied (in collaboration with Dr. M. Freedman of the School of Pharmacy, Toronto University).

4. To extend the technique to study protein-protein interactions initial studies are being carried out with α-lactaobumin (with Dr. J. A. Ferretti).

Honors and Awards: None

Publications: None

Serial No. 5.12
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: High-resolution Carbon-13 Fourier Transform NMR Spectroscopy at 55 MHz

Previous Serial Number: None

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Edwin D. Becker, Ph.D. (LPB:NIAMD) and Thomas C. Farrar (NBS)

Cooperating Units: Laboratory for Physical Biology, NIAMD

Man Years:

Total:	.40
Professional:	.40
Other:	.00

Project Description:

Objectives:

To develop equipment and methods to obtain natural abundance carbon-13 NMR spectra at 55 MHz (ca. 51000 gauss)

Methods:

Equipment is being constructed and a computer is being purchased to permit recording the free induction decay in an NMR experiment and then taking the Fourier transform after sufficient signal enchancement.

Major Findings:

Preliminary experiments indicate that a 100-fold enhancement in the carbon-13 signal to noise ratio over conventional techniques for a given observation time is possible.

Significance to Biomedical Research:

Carbon-13 NMR provides a sensitive technique to correlate molecular changes with biological activity such as changes in conformation and active binding sites in enzymes.

Honors and Awards: None

Publications:

Becker, Edwin D., Ferretti, James A., and Farrar, Thomas C.: Driven equilibrium Fourier transform spectroscopy. A new method for Nuclear Magnetic Resonance signal enchancement. Journal of the American Chemical Society 91, 7784-7785, 1969.

Serial No. 5.13
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Rate Studies by Nuclear Magnetic Resonance Spectroscopy

Previous Serial Number: None

Principal Investigator: Stephen R. Heller, Ph.D.

Other Investigators: James A. Ferretti, Ph.D., and Mildred McNeel

Cooperating Units: None

Man Years:

Total:	.40
Professional:	.40
Other:	.00

Project Description:

Objectives:

By a very detailed analysis, to determine the limits of the NMR technique in rate studies and evaluation of thermodynamic parameters of activation.

Methods:

Use of programs written for NMR exchange in conjunction with MODELAIDE and the IBM 2250 graphic display.

Major Findings:

NMR spectroscopy is useful for determining rate information over a very narrow temperature range. Usually the energy and ΔF of activation are obtainable within 20%. However, there is generally considerable error associated with values obtained for the other thermodynamic parameters. The results of this work place considerable doubt on the reliability of much of the data in the literature obtained by NMR.

Significance to Biomedical Research:

No direct significance.

Honors and Awards: None

Publications: None

Serial No. 5.14
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Applications of Walsh Transforms in Nuclear Magnetic Resonance Spectroscopy

Previous Serial Number: None

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: George Weiss, Ph. D. and Ray Mejia

Cooperating Units: Laboratory for Applied Studies, DCRT

Man Years:

Total:	.10
Professional:	.10
Other:	.00

Project Description:

Objectives:

This project is an exploratory one to see whether Walsh transforms can be used in NMR in the same way as the usual Fourier transform techniques.

Methods:

In the initial phase, programs will be written to do the logical arithmetic as well as further processing of Walsh transformed data.

Major Findings:

Project just begun.

Significance to Biomedical Research:

Although this project has no direct biomedical significance, it is of considerable potential interest for the processing of certain types of biological data. The Walsh transform can be performed considerably more quickly than the Fourier transform. Hence there are

potential applications in the real time processing of EEG or EKG data.

Honors and Awards: None

Publications:None

Serial No. 5.15
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Models for the Transport Properties of Membranes

Previous Serial Number: None

Principal Investigator: Michael C. Mackey, Ph.D.

Other Investigators: Mildred McNeel

Cooperating Units: None

Man Years:

Total:	1.0
Professional:	1.0
Other:	0.0

Project Description:

Objectives:

To develop a theory of the electrical properties of membranes based on the ideas and techniques of kinetic theory.

Methods:

The methods common to statistical mechanics and the kinetic theory of gases have been used.

Major Findings:

The kinetic properties of the membrane have been shown to depend on the force law between the ion and membrane. Detailed properties predicted by this model, such as chord conductance, have been shown to be in agreement with some experimentally observed features of the electrical properties of membranes.

Significance to Biomedical Research:

The transport of ions across membranes is a major biological phenomenon, and one that has challenged theorists for many years. The

present approach has promise as a basic physical model rather than a phenomenological one.

Honors and Awards: None

Publications: None

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT 6
Serial Number

2. HEURISTICS LABORATORY

3. James R. Slagle
Chief

I. OVERVIEW

The Heuristics Laboratory is fundamentally a research unit, concerned with the application of computers to very complex problems which require for their solution that the computer exhibit behavior which would generally be regarded as intelligent. The development of methods or strategies (i.e., heuristic programs) for use in such computer systems results in tools which can be applied to current difficult problems, including some in biomedical fields. From a long-term viewpoint, however, such tools are to be only parts of more extensive application packages.

More specifically, since its establishment the Laboratory has been concerned with the development of techniques and programs to perform automatic deductive question-answering and problem solving. Subsequently, work in the domains of automatic pattern recognition and automatic encoding of chemical information has been undertaken. In addition, the personnel in the Laboratory provide the Division with a source of expertise in areas related to heuristic programming, including use of computer languages for non-numeric programming and use of on-line, time-shared, interactive computer systems, as well as a background of knowledge in many areas of mathematics, such as formal logic.

New concepts and techniques developed by the Laboratory in the course of its research are thoroughly analyzed, both theoretically and, whenever desirable, by means of experiments with computer programs. It is a policy of the Laboratory that programs are written in widely-used computer languages (mainly LISP and FORTRAN), so that other installations may make use of our products. Virtually all of the Laboratory's computing is done on a PDP-10 equipped with a time-sharing monitor and special-purpose display hardware. This system is ideal because it meets the requirements for interaction and unusual resources that are characteristic of the experimental programs under development.

II. SUMMARY OF ACCOMPLISHMENTS

The research performed in the Heuristics Laboratory during the past year has led to significant progress. As is evident from the introductory discussion, progress in the development of heuristic programs (which is closely allied to the branch of computer science called Artificial Intelligence) is often not measured in immediate practical applications, but rather in terms of successful solutions of some of the difficulties which must all be overcome before

truly complex applications become a reality. Thus, for example, a more efficient deductive algorithm brings us a step closer to a practical deductive information-handling application..

The individual progress reports discuss the work of the Laboratory categorized in three more or less distinct areas. Applications of the research in each area to practical problems either have been or could be made today. Yet, in one sense the distinctions are artificial, since the complex systems which are the long-range goal of the Laboratory will combine results from more than one, if not all of these areas and others besides.

One of the characteristics of all the areas discussed is the necessity for handling large volumes of data. These quantities of data arise not only from large data bases, but from partial results and conclusions generated by the programs operating on that data. Thus, a question-answering program handling any reasonable query will not only draw from great amounts of information, but will generate many chains of inference in the course of arriving at a conclusion. For this reason, work in artificial intelligence has always needed to concentrate both on effective and efficient strategies and on compact, productive representations or models for data and the structures created while manipulating the data. Unfortunately, specific study of the consequences of large volumes of data has tended to lag behind the development of useful algorithms and heuristics.

A. The oldest area of research in the Laboratory, automatic question-answering (DCRT 6.1), has continued to yield new and more efficient strategies, as well as refinements and adaptations of old strategies which deduce consequences or "prove theorems" from sets of axioms or hypotheses. Looking to the future, work has been done to develop ways by which a program can improve its own strategy, and thereby "learn" to perform better as the result of experience.

This year has seen the start of the development of the so-called Programming Assistant, our first intensive effort to incorporate deductive strategies into a truly large package. As expected, this effort has revealed many new difficulties. Representation of even the moderately large amounts of data used so far has proved to be a sensitive task, if the relevant facts for any specific query are to be accessed quickly and without side effects from related but non-relevant data. In order to prevent costly regeneration of conclusions, means must be found to save and reuse results that are obtained.

Originally it had been thought that the Laboratory's venture into large-scale question-answering systems would involve a general-purpose program using a biomedical data base. However, after a thorough review of possible biomedical applications, we concluded that we did not have sufficient general knowledge about large-scale systems. There are enough unresolved difficulties associated with the handling of large volumes of data to prevent us from dealing with them effectively while at the same time coping with the specific complications introduced by a particular application. The choice of an in-house data base well understood by the investigators frees us to concentrate almost exclusively on the general problems of sheer size. Every effort is being made to keep the developing system general-purpose, with the realization that any finished application will have to be augmented by special-purpose heuristics.

B. Efforts in the area of automatic pattern recognition (DCRT 6.3) have continued to result in the development and analysis of new techniques. Here also the complications inherent in processing large amounts of data have been given increasing attention. Sequential pattern recognition, used when it is not feasible to test all features of the input data for reasons such as complexity or side-effects, has been studied, and new methods from other areas of heuristic programming have been applied to this process. In particular, a way has been found to express sequential pattern recognition in a form which allows techniques for searching trees generated in automatic game-playing to be used to decide which feature to evaluate next in the sequence.

The application of information-theoretic ideas to the problem of determining important features for evaluation has also been investigated, yielding further means for achieving effective pattern recognition in the context of large quantities of data.

C. This year a new project that is somewhat of a departure from the Laboratory's previous activities was begun. This work (DCRT 6.4) is concerned with automatic encoding and processing of chemical information, a necessary first step in building a practical, convenient, computer-based chemical information system. Thus, the main emphasis here is upon the use of a concise notation for representing large amounts of data, so that they can be dealt with efficiently. One of the many existing chemical notation systems has been chosen and programs have been developed to translate input into this notation and to manipulate information stored in this way. Large data bases already exist which could be accessed by the programs under development, suggesting that a large-scale application of Laboratory work may well come soon in this area. The existence of a chemical information system could then form the basis of attempts to automate various tasks of the chemist, such as the synthesis of compounds. It is expected that techniques developed for other automatic problem-solving systems would prove to be useful in such future research.

D. The personnel of the Heuristics Laboratory have also contributed to the Division in other ways. They have given seminars about their work and about the mathematical and programming tools which they employ. A course, "Symbolic Logic and Its Application to Artificial Intelligence," was given by two Laboratory members under the auspices of the Computer Center Branch. Members have provided individual consultation on other projects both inside and outside the Division. Finally, due to our heavy use of the PDP-10, the Laboratory has contributed to the on-going development of this computer system. We greatly appreciate the efforts of the PDP-10 Systems Group of the Computer Center Branch in maintaining this facility which is of such importance to us.

Serial No. DCRT 6.1
1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Automatic Question-Answering

Previous Serial Number: Same

Principal Investigator: James R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D., John K. Dixon, Carl Farrell, Deena A. Koniver, Richard C. T. Lee, Ph.D., Lewis M. Norton, Ph.D.

Cooperating Units: None

Man Years

Total: 5.6
Professional: 4.9
Other: 0.7

Project Description:

Objectives:

1. To develop, analyze, and improve strategies for answering questions and solving problems by computer.

2. To develop models and concepts for the representation of data, questions, and problems which allow efficient application of question-answering strategies.

3. To identify and solve difficulties arising from the use of question-answering strategies with large data bases.

4. To develop techniques for effective man-machine interaction to solve complex problems.

Methods Employed:

New strategies are programmed for the PDP-10 computer and extensive testing is carried out to determine performance characteristics. In general, some strategies are more effective for certain types of questions and problems than for others. Theorems are proved which make these distinctions precise and serve to further illuminate the properties of the strategies and to suggest new variations. Concurrently,

testing serves to suggest improvements in the design of data representations. Work on this project draws upon expertise in a wide variety of areas, including artificial intelligence, symbolic logic and set theory, probability and statistics, computational linguistics, and the use of time-shared computer systems.

Major Findings:

Investigations of deductive strategies have resulted in significant improvements, many of which are documented in the publications listed below. Techniques known as unit, input, and linear resolution and paramodulation have been analyzed. Resolution is a general method for manipulating facts (axioms). Paramodulation supplements resolution by providing special treatment for those facts concerning equality. In this spirit, various new techniques have been developed to give special treatment to other distinguished axioms, such as those of set theory. The SPECIALIZER, which is an automatic program writer, has been used to create some of the programs implementing these more efficient, special-purpose techniques.

One general-purpose program, MULTIPLE, has been the basis of a number of experiments. New, more discriminating methods for directing the course of a deduction have been developed, along with a method for automatic learning using regression analysis. Not surprisingly, the automatic learning is most successful when based on experience with a large number of problems.

An effort to incorporate the use of some of the deductive techniques mentioned above into a large data base system has resulted in the Programming Assistant, still under development, which is designed to aid in the creation and debugging of programs. Research into the identification and use of analogous previous results has been done for this system, and a natural-language user interface has been considered. Work with the Programming Assistant to date has resulted in two main conclusions: that moderately large collections of facts can indeed be handled efficiently, and that general-purpose strategies are not yet sufficient for a large-scale deductive system, but must be supplemented by context-dependent strategies.

One of the Laboratory's resolution programs has been made available to the general scientific community, and eight requests for it have been filled to date, indicating the interest in work on automatic question-answering.

Significance to Biomedical Research and the Program of the Division:

There is great potential for successful deductive question-answering strategies, not only for biomedical applications, but for many other areas to which computer science is being applied. A general-purpose, large-scale, interactive system for solving problems and answering questions would expand the volume of information immediately available to a research scientist and perform routine tasks for him, freeing him

to make more productive use of his time and other resources. In particular, the vast quantities of information concerning biomedical sciences makes a man-machine partnership the only means to successful conclusion of some research projects. Automatic question-answering would be a valuable addition to this partnership.

Proposed Course:

Research along the lines described above will continue, including both computer testing and theoretical analysis of new techniques. Adaptation of strategies to specific biomedical projects will be made when possible. Relative evaluations of existing strategies will be conducted. The Programming Assistant will be expanded in order to make further progress in the area of large data base systems.

Honors and Awards:

James R. Slagle, Ph.D., chosen one of the Ten Outstanding Young Men in America for 1969 by the United States Jaycess, partly in recognition of his work in heuristic programming and related areas of computer science.

Publications:

Chang, C.L.: The Unit Proof and the Input Proof in Theorem Proving. J. ACM. In press.

Author: Slagle, J.R.: Textbook of Artificial Intelligence. The Heuristic Programming Approach. McGraw-Hill. In press.

Slagle, J.R.: Interpolation Theorems for Resolution in Lower Predicate Calculus. J. ACM. In press.

Slagle, J.R., Chang, C.L., and Lee, R.C.T.: A New Algorithm for Generating Prime Implicants. IEEE Trans. on Computers C-19: 304-310, 1970.

Slagle, J.R., and Dixon, J.K.: Experiments with Some Programs that Search Game Trees. J. ACM 16: 189-207, 1969.

Slagle, J.R., and Dixon, J.K.: Experiments with the M & N Tree-Searching Program. Comm. ACM 13: 147-154, 1970.

Slagle, J.R., and Koniver, D.A.: Finding Resolution Proofs and Using Duplicate Goals in AND/OR Trees. J. ACM. In press.

Serial No. DCRT 6.3
1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Automatic Pattern Recognition

Previous Serial Number: Same

Principal Investigator: James R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D.
Louis Hodes, Ph.D.
Richard C. T. Lee, Ph.D.

Cooperating Units: None

Man Years

Total:	2.3
Professional:	1.9
Other:	0.4

Project Description:

Objectives:

1. To develop new techniques for automatic recognition of patterns by computer in any type of data.

2. To solve specific pattern recognition problems which arise in the automatic question-answering project or in any other projects of the Laboratory.

Methods Employed:

Theoretical analysis of new algorithms accounts for the majority of the research on this project. Concepts and techniques have been taken from other areas of mathematics and applied to pattern recognition problems. Many algorithms have been programmed and tested using the PDP-10 computer and its associated graphical display capabilities.

Major Findings:

As a result of viewing sequential pattern recognition tasks as games against nature (chance), the many techniques for searching game trees

have been adapted either directly or by analogy for use in such tasks. Sequential pattern recognition is a widely used method of classification, whereby rather than testing all features of a pattern at once, one feature is selected and tested, the result of the test determining whether or not further tests are needed, and if so, which.

New programs for finding solutions to sets of inequalities have been developed and made available to the Division.

Criteria have been analyzed for distinguishing convexity in digital representations of images. Approximation by convex bounding polygons has been found to be a useful approach.

Application of information theory to pattern recognition problems involving large quantities of data has provided a criterion for selection of the variables which it would be most profitable to consider. These "information-rich" variables can then be chosen using dynamic programming.

Significance to Biomedical Research and the Program of the Division:

Pattern recognition is directly applicable to a large number of areas, such as image processing and data analysis. In particular, most diagnostic procedures conform to the model of sequential pattern recognition, making the application of a new body of techniques to this method an important advance. The work concerning criteria of convexity is valuable because of the possible distortion of properties such as convexity during the conversion of pictorial data to digital form.

Proposed Course:

Theoretical work will continue, with associated computer testing when appropriate. Techniques will be made available for specific research within the Division as they are developed. A systematic survey of pattern recognition techniques useful for medical diagnosis is being prepared. This presentation will be oriented to physicians rather than statisticians.

Honors and Awards: None

Publications:

Chang, C.L.: An Iterative Algorithm for Pattern Recognition. IEEE Trans. on Computers. In press.

Hodes, L.: The Logical Complexity of Geometric Properties in the Plane. J. ACM 17: 339-347, 1970.

Hodes, L.: A Programming System for the On-line Analysis of Biological Images. Comm. ACM. In press.

Hodes, L.: Discrete Approximation of Continuous Convex Blobs. SIAM J. App. Math. In press.

Serial No. DCRT 6.4
1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Chemical Notation and On-Line Information Systems

Previous Serial Number: None

Principal Investigators: Allen Chauvenet, Carl Farrell, Deena Koniver

Other Investigators: Richard Feldmann (CCB), George Miller (since 4/70)

Cooperating Units: None

Man Years

Total:	2.4
Professional:	2.3
Other:	0.1

Project Description:

Objectives:

The overall objective of this project is to enable a chemist to communicate easily with computer-based chemical information systems. An immediate goal has been the development of a computer program which can translate from connection tables to Wiswesser Line-Formula Notation (WLN). In support of the overall objective, an additional goal has been the investigation of the role of the WLN translator in information storage and retrieval applications.

Methods Employed:

A chemist (or user) may be pictured as one component of a large computer-based information system. Each component of such a system should describe or document chemical compounds in a manner most suitable for that component. At the same time it is required that each component be able to communicate effectively with the other components.

An Input Program was developed which employs the Rand Tablet/display on the PDP-10 to permit the chemist use of his "natural language" of structural diagrams. Compounds may be drawn by the user and are stored by the program in connection tables, which have been found to be the most suitable representation for the central component of the system. A program now under development will convert a compound from connection table form to Wiswesser Line Notation. WLN is a representation which

is frequently used to store information in large chemical information systems.

A search program has been developed which allows a user to examine a WLN data base for similar compounds. A structure generation program will derive structural diagrams from retrieved WLN strings and print the diagrams on the line printer. It is thus possible for a chemist to use this system to ask a structural question and receive a structural answer.

Major Findings:

No previous work had thoroughly studied the problem of automatically encoding compounds in WLN. It has now been demonstrated that this objective can be effectively met. An advantage of the WLN program is the confidence one can have in the output. Hand encoding is always subject to human error.

The programs have been demonstrated to several members of the chemical community. It has been suggested that the WLN search program would be best used as a preliminary filter for a detailed connection table search.

The generation of structures from WLN can also be done efficiently, thus raising the possibility of eliminating explicit connection table information from stored chemical records.

Significance to Biomedical Research and the Program of DCRT:

An on-line information system is a desirable, if not essential, tool to provide chemists with access to current literature and chemical information. A chemist would be most likely to personally <u>use</u> such a system if he could communicate with it via structural diagrams, allowing the system to translate between various representations used by its diverse components. Such a system would greatly assist the work of NCI, NLM, and other groups on the NIH campus.

Proposed Course of Project:

Development and expansion of existing programs will continue. The WLN encoder handles compounds containing not more than one non-benzene ring. Much of the work necessary for extension of this program to complex ring systems has been done or is presently under way. Further studies will be made in connection table or diagram generation from WLN.

It is intended to write programs to interface between the NIH system and existing Chemical Abstracts tapes and connection table search programs.

Study will be made of the general chemical notation problem with a view towards extending existing notations to include information of particular interest to physical or quantum chemists. It is an objective

of this project to allow all users a natural mode of expression for their chemical information needs.

Honors and Awards: None

Publications: None

July 1, 1969 through June 30, 1970

PHS-NIH
Division of Computer Research & Technology

Summary of Branch Activities — 1. DCRT-7

7. Data Management Branch — 3. Arnold W. Pratt, M. D.
Acting Branch Chief

I. SUMMARY

During F.Y. 1970 the Data Management Branch (DMB) made excellent progress on its mission to provide NIH with efficient, effective and easily accessible computer-based capabilities for data management and analysis. Its primary strategy remains the use whenever possible of standard programs and facilities to minimize time and cost in the development of systems and to allow DMB clients to handle the specific data management tasks inherent in their own work.

The great majority of those tasks continue to be creating, editing, updating of and reporting from files of well-defined, formatted data. During FY'70 the Design Methods Section, DMB, extended its Standard Update Facility, giving users directly the ability to specify basic types of input data edits. With this extension the Section can show the NIH staff (scientific and administrative) how to edit as well as create and update their own formatted files. The Section is also teaching the use of a more powerful query program by which those same NIH users can get information from their computerized files directly, without the need for a DCRT computer programmer to write a program for each query.

Because of the success of this approach in FY'70, the Branch was able to close down its punch card processing unit in the Westwood Building, without adding additional manpower requirements on the Institutes which it served. This move was significant for several reasons. It demonstrated that the basic philosophy of standard data management facilities could be implemented in the NIH environment. It provided those users with greater understanding and control of and responsibility for their own data management. (The NCI Japanese Migrant Study is an example of a project which might well have avoided the need for a man-year of DCRT analysis and programming if this capability had existed previously for NIH users.) The move implies that both the user and DCRT can concentrate on other "non-standard" tasks. This becomes increasingly important in the face of dwindling personnel ceilings for the Branch, the DCRT and the rest of NIH.

Such (currently) non-standard tasks appear among the selected individual project reports for this Branch which follow. The NICHD longitudinal study required special editing programs for its large variety of numeric data. The Interferon Information System (NIAID) makes use of the Conversational Programming System (CPS) to provide the NIAID investigators with immediate access both to their current research data and to the computational capabilities of CPS. The NCI Carcinogenesis Data System will require and the Design Method Section is currently planning another extension of the Standard Update Facility to handle needs for indexed sequential (ISAM) file processing.

The NIH needs cannot be met with formatted file processing techniques alone. Much of the information in which NIH deals is variable in length and content. A document or record frequently includes "free text" as well as well-defined formatted data. The DMB has a second set of standard facilities to deal with this kind of information. The facilities comprise an element-based information processing system, referred to as the Publication and Retrieval System (PRS). This system has "standard" applications, i.e., the preparation of compendia of abstracts, such as those published by NICHD. Several developmental projects are underway to test and demonstrate its applicability to new tasks, such as the processing of the "abstracts" which are part of contract or grant documents. The Pilot Data Management Project for NIGMS is one example of these efforts.

Finally, NIH like any large organization will always have needs which require the construction of specially designed systems to handle large volumes of transactions, to cope with complex organizational interfaces, or to meet other requirements beyond those with which the standard DMB software facilities can cope. During FY 1970 the Branch assisted the NIH Office of Financial Management, NIH, by creating a special system to handle the Indirect Cost Payment Pool for grants. The Branch also finished the analysis and design for a BEMT Information Reporting System. A more ambitious undertaking was the creation of a new Project Accounting System for DCRT itself. The system is in many ways the most technically advanced administrative data processing system at NIH. It was a developmental effort within DCRT, the technical and managerial experience from which will be transferrable to other NIH projects within the Branch in the future.

The Branch staff served as consultants and advisors to parts of NIH on problems extending beyond data management itself. Work with ODPPE and BEMT included studies of medical school financing, development of physician supply models and related studies of physician distribution. The Branch also works with other parts of DCRT on advanced technical development projects.

A. Organizational Developments

The details cited above should not obscure the main direction of the development of the Branch itself. During the year the Processing Methods Section and the Systems Application Section were transferred from the Branch to the Computer Center Branch, and at the end of FY 1970 the focus of the Branch is clearly Information System Development and not the maintenance or operation of such systems. With this focus came a necessary concommitant emphasis first on creation of systems which are directly usable and maintainable by the clients of the Branch and then on the training of clients to be capable and self-sufficient in the use of these systems.

In addition to developing systems, the Branch develops people. It is a measure of its success that its alumni have been hired by BEMT, NLM, and OFM, as well as by several firms in the private sector.

Serial No. DCRT 7.1
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Standard Update Facility

Previous Serial Number: None

Principal Investigator: John Emmett Ward

Other Investigators: Mary Sullivan, Louis Grosman

Cooperating Units: NIGMS, NIAMD, NINDS, NICHD, CC Blood Bank, BEMT, NCI

Man Years

Total:	5.7
Professional:	5.7
Others:	0

Project Description:

Objectives:

Provide a standardized tool for establishing and maintaining a computer file with a minimum of programming effort. Minimize the time and effort for converting users from manual and punch card systems to computer tape and disk systems. Align the use of this function with the existing Query system, thus providing the potential user with a means of querying and reporting on the data he has created. Additional objectives are ease of maintenance and customer independence. Built in error checks and adequate accompanying error message are provided with the update and include the following:

1. No duplicate records are allowed on the file.
2. No modifications to non existing file records are accepted.

Methods Employed:

The Standard Update is a single catalogued procedure with overall user's control options specified by an input control card. Control for each input transaction record is specified within the record by an action code (insert, modify or delete), file control number, and specifiers for position and length of the target field within the file record. Editing is handled by exits to a routine for comparison with user-specified edit mark for new or updated records.

Major Findings:

The facility has been successful for users in NIGMS, NIAMD, NINDS, NICHD, NCI, BEMT and the Clinical Center Blood Bank. The design of the system has proved acceptable for users having a minimum of previous computer experience.

Significance:

The programs required by this system are standard and have been comprehensively tested. The potential computer requirements for structured file users can be satisfied with attention to user essentials as opposed to programming and computer needs. The elapsed time from request for service to system implementation is reduced significantly along with the library of programs to support NIH research needs.

Proposed Course:

The system will be extruded to the Indexed Sequential Occur Methods (ISAM) for random access desk files.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.2
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Carcinogenesis Bioassay Data System Analysis

Previous Serial Number: None

Principal Investigator: Mary M. Sullivan

Other Investigators: Robert Drummond

Cooperating Units: Program and Data Analysis Unit Office of the Associate Director for Carcinogenesis, Etiology Area, National Cancer Institute

Man Years

Total:	.37
Professional:	.33
Others:	.04

Project Description:

Objectives:

Design a system with uniform requirements for gathering and processing standard compendia of information which are fundamental to carcinogenic studies. Design a system which is flexible enough to permit the collection and retrieval of information about unique test conditions, activities, and results. Design a system to provide a basic format to structure reporting of test results in all investigations. Design a system to facilitate contract administration and permit more satisfactory evaluation of experiments.

Methods Employed:

Standard Information Analytic techniques have been used.

Major Findings:

System analysis is complete and system specifications have been presented to the Program and Data Analysis Unit, NCI. Specifications for three major programs have been developed and programmer orientation has begun.

Significance to Bio-medical Research and the Program of the Institute:

The Carcinogenic Bioassay Data System is intended to facilitate analysis of experimental results on large numbers of animals and tests by imposing structure upon massive collections of data so that comparisons and correlations can be made upon experimental results within an investigation and among multiple investigations. It is expected that establishment of files will allow for more sophisticated analysis of experimental results.

Proposed Course:

Implementation of the basic system will occur during the next year.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.3
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Pilot Data Management System for NIGMS Grant Data

Previous Serial Number: None

Principal Investigator: Alan Kreger, DCRT

Cooperating Units: NIGMS

Man Years

Total: 0.1
Professional: 0.1
Others: 0.0

Project Description:

Objectives:

To develop a pilot system for the storage, retrieval and publication of NIGMS grant data. The data included will be both formatted and textual. It will include information from Grant Applications, Summary Sheets, Progress Reports and Reports of Expenditures. Quick access to user queries will be made by remote terminals.

Methods Employed:

DCRT's Publication and Retrieval System, (PRS).

Major Findings:

None to date.

Significance:

The capabilities of PRS to handle free text as well as formatted data should greatly increase the ability of program manages to maintain and analyze computerized files with a higher information content and shorter response time than the manual and automated systems currently in use at NIH.

Proposed Course:

A trial of the proposed system will be undertaken. Ten grants in the Trauma area will be used as the experimental data base.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.4
1. Data Management Branch
2. Design Methods Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Japanese Migrant Study (Shifts in Cancer risk for specific sites among Japanese migrants)

Previous Serial Number: None

Principal Investigator: Judy Mahaffey, DCRT

Other Investigators: James Gilliss, DCRT

Cooperating Units: Biometry Branch, Etiology Area, NCI, NIH

Man Years

Total: 1.8
Professional: 1.3
Others: .5

Project Description:

Objectives:

To do the systems design, programming, and documenting of a consolidated file for investigating factors which might account for the markedly higher lung cancer risk and moderately lower stomach cancer risk among Japanese migrants and their offspring when compared to the experiences prevailing in Japan.

Major Findings:

Although data collection by NCI had been going on for some time, limited statistical use was made of the data. Problems encountered in the initial DCRT systems analysis involved variances in coding structures, variances in the questions appearing on the questionnaire for gathering the data, multiple responses producing "illegal" coding structure, and inadequate data documentation. Therefore, to create a Standard file, new coding structures were devised to suffice for both present and future studies regardless of the geographical structures and underwent edit procedures to identify errors. Update procedures were produced to correct the erroneous data and/or add additional data. Control check points procedures have been implemented to insure greater accuracy.

Serial No. DCRT 7.5
1. Data Management Branch
2. Information Systems Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Element File Based Retrieval System

Previous Serial Number: DCRT 7.5

Principal Investigator: Hal Fredrickson

Other Investigators: Darlene Myer, Lawrence Weiner

Cooperating Units: None

Man Years (computed for the 12 month period)

Total: 1.25
Professional: 1.25
Others: 0.0

Project Description:

Objectives:

To provide NIH with the capability of storing and retrieving document text data from a user definable computer data base. This system must be "flexible" enough to encompass a wide variety of applications.

Methods Employed:

An "English-like" request language is used to enter the retrieval request from a remote computer terminal. The result of the requests are then retrieved to the terminal or the high-speed computer printer. The user has the option of defining the types of fields used in his "document" data bank and then defining his own symbolic name for operators, operands and variables.

The document file is a text oriented file that will allow up to 250 elements per document. The present maximum length of a document is 8,000 characters on magnetic tape or approximately 7,000 characters per element on the disk system.

A search can be performed on any text field(s) of the documents in one of 3 methods: 1) a character by character field match 2) free text search for word or word groups 3) free text "root word" search.

A sequential tape search is available for low usage files; this will accept request from a remote terminal, search the entire tape in the computer batch job stream, and then return the results to a remote terminal.

A direct access online search is being developed for high usage files. The cost for online storage may be relatively high but the computer cost per search will be lower than the sequential method. The online capability will allow more user interaction with his data bank and this will hopefully improve information recall. The user must specify which fields are to be used for retrieval--these fields are then inverted by key words and stored on three Index sequential disk files. The full text element fields are stored on variable length Index Sequential Files.

Major Findings:

A "flexible" user-oriented language for text retrieval has been developed for NIH. Current IBM OS/360 support for variable length direct access files is now satisfactory for the storage of our document files.

Significance to Biomedical Research:

The retrieval system will provide scientists and others at NIH with the facility to maintain and retrieve information describing completed work or work in progress on projects of many varieties.

Proposed Course:

Continue development of computer terminal support for the system and further develop the online data file capabilities.

Honors and Awards: None

Publications: None

Proposed Course:

In addition, the system affords the capability of incorporating into it additional sites of the body where cancer may strike (breast, prostate) when this data becomes available.

Significance:

The system, as designed and employed in processing the input data, is tailored to work in conjunction with the DCRT Library programs. Statistical Analysis in many cases can be satisfied by using TABLEMAKER or other generalized programs. Therefore, data analysis can be made by the National Cancer Institute Investigators independently of DCRT programmers and analysts. Adequate documentation has been provided to allow NCI to pursue independent statistical analysis.

Proposed Course:

The consolidated file and the file maintenance procedures, and their documentation will be turned over to the National Cancer Institute. Scientific findings and future publication will be through the planned efforts of the National Cancer Institute.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.6
1. Data Management Branch
2. Design Methods Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Longitudinal Studies Program of Human Physiology, Biochemistry, and Psychology (Aging Study): File Development Phase

Previous Serial Number: DCRT 7.6

Principal Investigators: Catherine W. Staneck, DCRT
Arthur H. Norris, NICHD

Other Investigators: DMB-Unit 7

Cooperating Units: Gerontology Research Center, NICHD

Man Years

Total: 2.5
Professional: .75
Others: 1.75

Project Description:

Objectives:

To facilitate the study of the effects of aging in a population of approximately 700 men, it was determined that a central masterfile containing the results of various physical, psychological and sociological tests should be established.

Methods Employed:

A series of Fortran and Assembler Language edit routines have been developed for use in conjunction with general purpose file maintenance routines currently available in the DCRT program libraries.

Major Findings:

Since the data for this file was originally located in approximately twenty-six different files, each with a unique format, it was impractical to develop, edit and update procedures on a field by field basis. In terms of the file development phase of this project, a major finding of the analysis has been the usefulness of producing mean values on numeric fields. As a part of the edit procedure, these

means are then used to identify individuals whose results are outside the range of two standard deviations from the mean.

Significance:

The development of a single masterfile from many different files has shown the advisability of using basic statistical measurements as edit tools. By calculating annual means by age group for various values, problems and inconsistencies in data collection have also been identified.

Proposed Course:

Eventually the editing and updating of the central file will be taken over by the Gerontology Research Center, NICHD, which is also handling the data collection. A program for reporting each patient's complete history of test results will be developed.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.7
1. Data Management Branch
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Interferon Information System

Previous Serial Number: None

Principal Investigator: Robert J. Drummond, Hector Motroni

Other Investigators: None

Cooperating Units: LVD/NIAID

Man Years

Total: .5
Professional: .5
Others: 0

Project Description:

Objectives:

To provide scientists immediate access to background data and results of previous experiments. To introduce research personnel to the computer as an analytical tool through use of an interactive system.

Methods Employed:

Data enters file through punched cards (batch processing). In this way file is maintained and updated with minimal time required of lab personnel. Most recent experiments (3 months) are maintained on a direct access file and are immediately accessible in the lab through a terminal (CPS). Older experiments are placed on tape in format identical to that of the current file. Thus researchers may analyze current experiments in the laboratory and receive immediate feedback on the terminal. Analysis of all past experiments may also be initiated through the terminal (using the past experiment tape) with results printed on the high speed printer in the computer center.

Major Findings:

The System has only recently become operational. Its effect on Interferon assay programs is not demonstrable.

Significance:

This project went beyond most, if not all, previous DCRT projects in support of laboratory research. The DCRT analytical systems analysis disclosed a new perspective in which the NIAID scientists could view about their data. The design of an integrated system to function through a terminal supported facility (CPS) is a demonstration that such facilities can be useful to the laboratory scientist.

Proposed Course:

Develop standard queries and statistical programs for further analysis of experimental results in order to aid in the planning of future experiments.

Honors and Awards: None

Publications: None

Serial No. DCRT 7.8
1. Data Management Branch
2. Design Methods Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1969 through June 30, 1970

Project Title: Project Accounting System

Previous Serial Number: None

Principal Investigator: Gerald D. Stoner

Other Investigators: Michael Crum, DCRT
William Jones, DCRT
Lanny Owens, DCRT
Louise Saulnier, DCRT
Kenneth Shapiro, DCRT
William Speary, DCRT

Cooperating Units: Project Control Office
Computer Center Branch
Financial Management Branch

Man Years

Total: 3.75
Professional: 3.5
Others: .25

Project Description:

Objectives:

The purpose of this project was to design and implement a new computer-based accounting system for DCRT. This accounting system should capture and pre-validate all input data to the system, maintain an accounting system master file, prepare necessary reports to management, interface with the Central Accounting System and make available to management data to be used for any of its special report requirements. Remote terminal input was desired for decentralized data input.

Methods Employed:

The system was making use of almost the full range of the DCRT computing facilities. All input data is either computer-generated or captured at tele-communications terminals, and kept in on-line files

until used by the update segment of the system. Error correction is also accomplished on-line. The system is dynamic, in that all the transaction files are automatically reset by the system. The design combines both Batch and On-line Processing, using PL/1, CPS and Wylbur.

Major Findings:

It was found that PL/1, CPS, and Wylbur can be used to good advantage to complement each other and that by using the Conversational Programming System to collect and pre-validate data, two major advantages are enjoyed: 1) data can be easily and efficiently entered by non-data processing personnel, and 2) the data going into the system update is very clean because of the input validation.

The system became operational April 1, 1970.

Significance:

The system is the first at NIH which serves the needs of an entire organization with remote terminal entry of data, on-line/real-time validation to reduce input errors, on-line terminal correction capability for errors which slip past initial validation and are rejected at times of master file update. Development of the system demonstrated drastically the often discussed but seldom attained benefits of a careful, recursive system analysis and design phases prior to implementation as such it can serve as a bench mark for other system development projects at NIH.

Proposed Course:

After the system "undergoes" operational testing and documentation, it will be turned over to the Systems Application Section, CCB and the Administrative Office, DCRT for recurring operation, maintenance and monitoring.

Methods Employed:

Standard data and file definition techniques.

Honors and Awards: None

Publications: None

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY
Report of Program Activities
July 1, 1970 through June 30, 1971.

ANNUAL REPORT

TABLE OF CONTENTS

Page

Page

Page

DCRT Annual Report
July 1, 1970 through June 30, 1971
Office of the Director

Introduction

Last year the Director's summary emphasized the internal functional overview of the activities in the Division. This year the focus is on the external program interfaces, showing how DCRT relates to, supports and complements other NIH activities.

The mission of NIH in biomedicine includes the performance at NIH of clinical and nonclinical research and the care of patients during clinical research. It also involves leadership and support outside NIH for such research and for health manpower education and the effective communication of medical knowledge. Finally, it entails the administrative management activities which are inherent in any enterprise as large as NIH. It is those NIH activities which are the framework for the programmatic overview below.

Individual Laboratory, Branch and project reports must, as always, "speak for themselves," since a project in a specific Laboratory or Branch often spans categorical boundaries, functional or programmatic, and incorporates more than one of the Division's basic disciplines: computer/information science, engineering, and mathematics.

Review of Functional Categories

Before concentrating on the NIH programmatic array, we ought at least to review the categories in the DCRT functional array.

I. Research
   A. "Intensive," i.e., directed toward computer and information science.
   B. "Extensive," i.e., directed toward biomedical science and practice.

II. Development of systems and components
   A. Software only
   B. Hardware only
   C. Joint software and hardware

III. Provision (development and operation) of computing services and facilities

IV. Provision of educational and informational resources

V. Provision of advice, assistance and consultation

VI. Development of people

Clinical Research

Computers at NIH and elsewhere have not yet made a profound or widespread contribution in clinical research. The reasons are easy to see and hard to change in a short time. Computers have been most useful for the processing of large amounts of well-defined information. The information must have clear meaning, be in machine processable form, and require more than trivial processing before a computer is obviously useful. Most clinical information exists on paper, defined only in the context of the medical record of a specific patient or in an investigator's notebook. Because of this and because of the usual necessary focus of the clinical investigator on a single phenomenon or problem in a limited number of patients, the kind of processing he considers doing is usually too small to make the creation of a machine processable data base and a set of computer programs feasible in terms of his experience and expectations.

Thus, instances where DCRT and the NIH clinical scientists are using computers are those where machine processing has appeared attractive or unavoidable or both. For example, cardiac function studies involve transducer signals of blood pressure and ECG activity which were being captured anyway and could be readily processed by computer once the intellectual and technical hurdles were lowered by collaboration between the NHLI and DCRT staff. In another case, the NCI Surgery Branch became aware of an existing set of programs which could help it analyze clinical data about the short and long term response to treatment of carcinoma of the cervix and of the head and neck. They then had the incentive to review charts on those patients and create a processable data base covering several hundreds of patients.

In the long run the effective and efficient use of computers in clinical research will depend upon the clinical scientist's increased awareness of the instances where computers can be useful. More than a knowledge of computers and programming languages, this involves an understanding of the principles of data management and analysis and a sense of taste which allows the scientist to match the problem against the methods for information processing and both against the costs. This kind of omniscience is rare and, for the present, progress will have to be made largely by DCRT offering an opportunity and supporting those clinicians who are willing to explore it.

Clinical Care

Clinical Research and Clinical Care are more intimately connected at NIH than in most hospitals. The NIH clinical mission is research. The emphasis in clinical care is, however, always on an individual patient and his problems within the immediate time frame dictated by them.

The divisions of labor and knowledge in a modern medical center result in greater need and opportunity for computing in particular departments. Computers are used or under consideration in the NIH Clinical Center by its blood bank, clinical laboratories, medical record and nuclear medicine

departments, nursing services, pharmacy, and various administrative services. The information content in these areas tends to be highly structured and formatted and its processing more regularized, if only because other physicians have become dependent upon each as a reliable "service."

The primary purpose in each case is the improvement of the particular operational activity, but the effort can provide a transferable product (technology) or publishable results (research). For example, the joint effort of DCRT and Department of Pathological Anatomy created programs for storage and retrieval of SNOP-coded pathological diagnoses which have been picked up by pathologists at other medical centers. The basic linguistic problems of analyzing and encoding medical English are a research focus within the DCRT.

What remains is "putting it all together," so that computing supports the physician himself in the act of caring for specific patients. Some work has been done on specific programs for assistance to clinicians in their diagnostic and therapeutic activities. Like most "total systems," the approach has to be evolutionary and will be based on initial success in specific medical subspecialties. The conceptual design of an overall clinical care support system is a worthwhile intellectual exercise, but engineering design and development of such a system is an entirely separate matter and requires experience which can be only gained in separate activities. A joint DCRT/Clinical Center Coordinating Committee now meets monthly to review progress and plans. Initiative and project leadership comes from the Institute clinicians, the DCRT and the Clinical Center.

Two things are essential to long term progress for the effective application of computers to clinical care. One is an adequate "systems base," i.e., the appropriate computers organized under a good operating system, designed to support the needs of a clinical environment. The other is a nucleus of competent physicians who have enough realistic insight into the potential benefits of computers in clinical medicine to work with analysts and programmers in designing and developing applications to run on the systems.

DCRT and NIH are already fortunate in having the technical competence to create or acquire and then to maintain and improve the systems base. During the forthcoming year (FY 1972), DCRT in conjunction with professional staff of the Clinical Center and several Institutes will offer an elective on "Computers in Clinical Medicine" to a small group of medical students. This will cover both clinical care and research. It will help to build the requisite nucleus of physicians both for NIH and elsewhere.

<u>Nonclinical Research</u>

This category includes work in laboratories and those epidemiological studies which fall into the province of public health rather than clinical research. It is the area where DCRT has, to date, made the greatest impact in its mission to incorporate mathematics, engineering, and computer research and technology into modern biomedicine.

The work of the Physical Sciences Laboratory supports and complements many laboratory activities at NIH, particularly in development of the theoretical basis of research. The work covers a broad range of biochemical and biophysical phenomena as a reflection of the diverse interests of the members of that laboratory. To extend the range of interaction of his staff and of others in DCRT with biomedical scientists throughout NIH, Dr. Weiss proposed and developed a roster of DCRT Scientific Consultants. The list includes applied mathematics and statistical groups within the Laboratory of Applied Studies, which has its own on-going collaborative work with laboratory research throughout NIH, and members of the Heuristics Laboratory, which seeks such collaboration.

It also includes members of the Computer Systems Laboratory, which has profoundly influenced computing in the NIH laboratories by working with NIDR, NIAMD, NHLI, NICHD, and NIMH in the specification and development of computer systems for direct connection to laboratory experiments and apparatus in those Institutes. Without this resident CSL expertise it is improbable that the NIH could be in its current leadership position in the application of computers in direct support of laboratory research.

Finally, the Data Management Branch has created computer programs and systems of programs for many laboratory scientists. These run on the NIH central facilities provided by the Computer Center Branch. The reorganization of the Data Management Branch under its new Chief, Mr. Emmett Ward, augurs particularly good prospects for management and analysis of research data at NIH. Its Software Support Section, Scientific Application Section, and Math/Stat Programming Section all have acquired new strength, and during the coming year we expect striking progress.

Communication of Biomedical Knowledge

The NIH mission to support the communication of biomedical knowledge has been seen largely in the National Library of Medicine. The DCRT work in support of biomedical communications has been limited to some work in support of the operations of the NIH library, to the development of MARC, a prototype touch-tone data entry and audio-response communication system, and to the development of systems to create and query files of information about chemical compounds and their attributes.

One of the chemistry systems is unique in its ability to enter information as a chemical graph through a Rand Tablet and then to store information internally as either connection tables or Wiswesser Line Notation format to present information in graphic form. Two others provide for key word searching of CAS-CBAC literature citation files: one responds on-line, but provides only the identification numbers of the relevant citations; the other provides a more complex search and complete output by batch serial searching.

The DCRT Library and the DCRT Scientific and Technical Information Office have a unique responsibility for communication in that subset of knowledge relating computing, mathematics, engineering and related disciplines to

biomedicine. They also constitute a basis for possible future research and development in information retrieval and communication techniques. At present, however, they are limited almost entirely to maintaining existing operations and coping with requests.

National Leadership and Support of Research

This NIH category overlaps both the performance of research and also some administrative tasks which are part of the grants and contract operations. DCRT has no programmatic mission in the leadership and support of biomedical computing, except by the influence of its own research and technology and staff. DCRT has exerted leadership as advisor and consultant to NIH grant and contract programs in which computing is a prominent component.

Conversely, there has been little use of the basic DCRT systematic disciplines (mathematics, engineering and computer/information science) for managerial functions throughout NIH. This is not surprising since to date the application of these systematic disciplines under the various banners of management science, operations research, systems analysis, etc., has not appeared useful to many managers at NIH. Furthermore, although the DCRT embodies the disciplines it is not yet viewed by managers at NIH as a sourch of analytic talent for their work. This area will probably evolve gradually over the next few years as both DCRT and the NIH program leaders accumulate experience.

Health Manpower Education

The Bureau of Health Manpower Education is predominant here, and the DCRT has no programmatic responsibilities. Its only involvement to date has been to provide computer training for health professionals working at NIH, some of whom make use of it later in positions of responsibility outside NIH. Medical students have sought out DCRT informally for their electives, and the formal elective, Computers In Clinical Medicine, should be a useful addition to what is at present a limited educational experience in computing for medical students across the Nation.

Administrative Management

This category includes financial management, personnel management, grants management, supply management, property management, contracts management, and the management of administrative and engineering services. These occur at the NIH level, at the Bureau level or at an Institute or Division level. In the latter, they overlap some of the operational activities of grants and contract programs and also occasionally with "program analysis" activities. This category comprises the largest use of computer time at NIH. This is not surprising since it comprises virtually the total use in most agencies of the Federal Government.

The DCRT involvement is, in general, inversely proportional to the size of each specific activity. The DRG IMPAC system is largely independent except for its reliance upon the DCRT Computer Center machines and some attendant

advice about the operating systems under which they run. Other central NIH administration data processing activities regularly make use of DCRT key punching and recurring production operations, and depend to varying degrees upon the Division's programming services. The Institutes and Divisions are, in the main, much more dependent upon DCRT for analysis and design of their data processing systems as well as the development of systems of programs which perform the processing tasks. The Data Management Branch has given increasing leadership over the last two years both in the creation of the systems and in the development of mature independent users of those systems once created. It is probably no overstatement to say that the level of sophistication among users of administrative data processing at NIH is now as high or higher than any other part of the Federal Government. Still there is room for improvement.

What it all means

In summary then, computing in general, and DCRT in particular, shows an extensive involvement throughout the laboratories and clinics and the administrative management of NIH. DCRT involvement in biomedical communications is limited to some circumscribed areas; these complement the activities of the National Library of Medicine and support those of the NIH Library. The involvement of DCRT staff and their disciplines in the NIH managerial responsibilities for leadership and support of research, training and health manpower education is still maturing.

The accomplishments or lack thereof in all these areas are dependent upon the usual factors:

1. A recognized and defined objective (need or opportunity)

2. An adequate understanding of, or successful commitment to discover a way to advance toward the objective

3. Adequate resources (technical, intellectual and managerial) to move toward it, in a sustained way

In each of its research and developmental activities under the NIH programmatic array, the DCRT almost always shares the responsibility for these factors with some other part of the NIH. The DCRT responds to specific needs recognized by others; it takes initiative in areas where it perceives present opportunity or future needs. But the success of the endeavor depends upon the performance of people outside as well as inside DCRT. This is a major challenge inherent to the DCRT mission.

This dependence does not apply in the same way for those DCRT functions which do not appear in this year's discussion under the NIH programmatic array, i.e., provision of basic central computer services and facilities in support of the entire range of NIH activities, "intensive" research on basic problems of computer and information science, and the development of people whose ability to use computers effectively and efficiently advances the NIH mission. Examples of these internal DCRT functional categories may be seen in the Laboratory and Branch summaries which follow.

Serial No. DCRT 1.1
1. Office of the Director
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Visual and Biological Shape

Principal Investigator: Harry Blum

Previous Serial Number: None

Other Investigators: Virgil Carlson (NHLP)

Cooperating Unit: NIMH, Laboratory of Psychology, Section on Perception

Man Years:

Total:	0.9
Professional:	0.9
Other:	0.0

Project Description:

Objectives:

The overall objective of this project is to develop a formal description for biological shapes and apply it to the development of a psychology of shape and an elaboration of shape processes in visual physiology.

Methods Employed:

Develop and formalize a geometry of shape using growth processes. Explore implications of such a geometry to the visual process. Review relevant areas of experimental work in psychology and physiology. Design new experiments aimed at validating or refuting theoretical work.

Major Findings:

A major theorem concerning flexural objects was proved. Using it, a number of shape properties of flexural objects were analyzed. The experimental aspects of the

theory were investigated and a set of psychophysical experiments on shape vision were designed. Preparations for performing these experiments have been started.

Significance to Biomedical Research and the Program of DCRT:

The theory contributes a fundamental insight into several biological areas. First, it proposes a morphological base for description of cells, organs and organisms. This will be needed for taxonomic and diagnostic purposes, both human and automatic. Second, it opens up the area of shape perception which is needed for teaching and diagnostic purposes. Third, it proposes new principles of central nervous system organization which may be important to the neurophysiology of vision and to general understanding of the CNS. This extension encompasses motor, as well as sensory organization. Fourth, it has implications in embryonic, and later, growth and development since the theoretical principles apply to the efficient coding of biological shape. The visual and the growth coding are shown to be inversions equally well treated. Thus, the work, if correct, has potential for opening up new and important areas of biology which have been resistant to investigation before.

Proposed Course:

Theoretical work will continue. Experimental work on shape psychophysics will continue. It is hoped that experiments in visual physiology will be started.

Honors and Awards: None

Publications: None

Serial No. DCRT. 1.2
1. Office of the Director
2. Budget, Personnel, Administrative Office.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Administrative Information Retrieval System

Previous Serial Number: None

Principal Investigator: Michael A. Reed

Other Investigators: Hal Fredrickson
Darlene Myer

Cooperating Units: None

Man Years:

Total:	1.75
Professional:	1.25
Others:	.25

Project Description:

Objectives:

To provide a system whereby the administrative/management personnel may collect, maintain and search files of administrative information in a manner easily understandable and operable by a non-computer oriented person.

Methods Employed:

An input program which, in effect, interacts with the user, has been written in CPS language. This program assists in the collection and modification of a file of administrative information. The program references a prompt file which requests the user for information desired on an administrative subject (such as personnel, contracts, space information, etc.) and will, when the data is entered, provide edit checks on the data.

JCL programs have been written to 1) convert the collected information to a format acceptable for retrieval; 2) merge the collected information to a master file; 3) list the master file; 4) delete records from the master file. These programs were written in such a manner that the modifications necessary for running are minor and easily accomplished by a non-programmer.

The retrieval phase of this system is accomplished through the use of a previously written program discussed in project No. 1.2 Essentially this program uses an "English-like" request language and thereby allows the non-programmer to specify his own requests. The program also allows for varying formats, certain mathematical operations, and simultaneous multiple requests.

Major Findings:

The computer and programs to run it do not necessarily have to be operated by computer type personnel. With some effort and through close cooperation with administrative managers, a system has been developed which comes one step closer to making the computer more usable by a greater variety of people. This system is fully oriented to non-computer personnel.

Significance to Biomedical Research:

This system is designed for use as a tool by the NIH program manager in the conduct of his day to day business. As such its significance to biomedicine is found in the assistance it lends the manager in organizing and controlling the vast amount of data necessary for efficient management of program activities.

Proposed Course:

Some programming is still necessary before the system will be completely operational. The primary course after debugging will be minor modifications to set up the system for the information needs of the various users. At present this system is being used in the collection, etc. of information on personnel and other administrative type data. Future expansion is planned in other areas of administrative concern.

Serial No. 1.3
1. Office of the Director
2.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Acquisition of Physician Orders

Previous Serial Number: None

Principal Investigator: Dale N. Gerding, M.D.

Other Investigators: None

Cooperating Units: National Institute of Arthritis and Metabolic Disease, Arthritis and Rheumatism Branch.
Department of Nursing, Clinical Center.

Man Years:

Total:	1.5
Professional:	1.5
Others:	0.0

Project Description:

Objectives:

The ultimate objective of this project is to acquire in machine processable form those physician orders which determine and initiate essentially all of the diagnostic and therapeutic activities of the hospitalized patient. In particular, those orders pertaining to drug therapy are of interest. Such information provides the basis for ultimate analysis of drug efficacy, toxicity, and interraction, as well as the eventual reduction of drug administration errors.

Methods Employed:

A DCRT supported public text editing system called WYLBUR has been the means used to acquire the basic data. Data is entered via computer typewriter terminal at the nursing unit, and is transmitted by telephone to the DCRT Central Computing Facility.

Major Findings:

The system has been in full operation on a single nursing unit for over two months. Nursing response to the system has been unanimously favorable. At least six documents of clinical use to nurses are being generated daily from input data. Achieving compatability with nursing has been the initial objective of the system and this has been accomplished.

Proposed Course:

Additional programming is underway to produce patient files of ordered drugs. From such files a system will be developed for the recording of actual drugs administered to the patient. This information will then be compared to clinical and laboratory data in the evaluation of drug toxicity and efficacy.

Honors and Awards: None

Publications: None

PUBLIC HEALTH SERVICE-NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

1. DCRT-2
Serial Number

2. COMPUTER CENTER BRANCH

3. J. D. Naughton
Branch Chief

I. SUMMARY

The Computer Center Branch designs, plans, implements and maintains a large general-purpose computer utility to meet most effectively the dynamic and diverse requirements of both N.I.H. research investigators and managers in the support of modern medicine. This charge includes the development of unique system facilities that will bring the computer to bear on problems at every level of research and in many locations. The core of this computer facility is a network of computers and remotely located terminals, which, by means of modern communications techniques, makes the power of the computer immediately available in all laboratories and offices throughout N.I.H. An inherent responsibility of the utility is the continued research into extending the network into the research environment and adapting to the constant impact of new knowledge and program direction.

A full spectrum of computing power is provided to research personnel.These facilities include conversational programming, graphics, text editing, remote job entry and batch processing. Large systems as well as mini-computers and terminals are lashed together providing capability at many levels. Research into the computer and information sciences coupled with the N.I.H. medical investigators' needs and cooperation will bring computers closer to the research environment where they can perform most effectively in attacking the complex problems of modern medicine.

The medical research programs of N.I.H. require the most powerful and flexible of computer services and tools available today. The computer network provided must have a distributive power that is easily accessible on demand to scientists in the laboratory itself. The goal is to mold, polish and, ingeneral enhance the computer into a complete tool for medical research and its administrative support. New areas of computer applications will be sought out continuously, in conjunction with a comprehensive program of educating new and old computer users in how to use computers most effectively in their domains.

In response to the ever increasing demand for computational support, by both the N.I.H. scientist and researcher, the N.I.H. computer utility has expanded greatly in size and function. The computer network grew from 10 to 18 remote batch job entry terminals, from 100 to over 250 interactive typewriter terminals, and from 3 to 4 central processors. The overall workload of the Computer Center has increased 112% to over 87,000 jobs per month during the year. Terminal sessions increased from 400 to over 1000 per day. Rates during this period were reduced 8%. The acceptance and growth of terminals drastically changed the profile of the average users job and his work habits, generating a need for powerful on-line daytime processing, that was answered partially by the additional central processing unit. The change in work profile had a dramatic effect upon the amount of disk storage shared by the CPU's of the network. The large dependance upon on-line data bases has swelled the amount of shared disk space to over one billion bytes of storage to accomodate the users program needs.

The popularity and utility of the interactive terminal systems, WYLBUR and CPS, is reflected in the number of terminals installed and by the amount of work generated by terminals. Over 50% of the current work load is now generated by terminals. The use of computers by means of terminals spreads into new areas continuously. Notable examples of use are the Blood Donor System of the Clinical Center's Blood Bank which uses both WYLBUR and CPS to find compatible blood donors, and the Cancer Institute's Drug Data Sheet program which depends upon CPS to obtain dosage information, toxicology and the mechanism of action of drugs in its data base. Wylbur has also been used for the rapid preparation of documents needed by the administrative processes. Demand for these services has been such that the Computer Center has, in the last year expanded the number of telephone lines available for these services to over 150 and is planning to expand the service to 180 lines in the upcoming months.

The molding of our computer operating systems to remain sensitive and responsive to the research community's needs continues. The latest pertinent computer science techniques are employed to maintain or reduce response time while servicing an ever-increasing number of users. New techniques are continuously studied and installed to increase the reliability and thruput of productive systems. A major modification to our operating system that allows all of the Center's computers to share a common work queue, providing dynamic load balancing, was successfully installed. The user benefits from the new features added and the continued high level of service.

INTERFACE, a series of technical notes, continues to keep the Computer Center users informed of the latest in computing and Computer Center activities. The Computer Centers' Users Guide under went its third complete update. The effort spent supporting and keeping the Computer Center's users up to date also increases. The success of work in this area is attested to by the wide acceptance of this group by users of the Computer Center.

The implementation of a unique software support system for an IBM 1288 Optical Page Reader has opened new avenues of data entry to the Central Facility. The general purpose nature of the system enables it to be responsive to rapid changes in data collection requirements. Merely by describing documents in terms of a set of parameters the user can go from document to magnetic tape with a minimum of effort and time. The system makes full use of the pencil and the typewriter as effective data entry mechanisms.

Investigation into the processing of on-line files continues. One file query system was implemented, used, and discarded as the system proved to be restrictive and the attributes were such that the system did not meet the needs of any large group of users. Currently under investigation is a new system for on-line file processing which , if successful, will allow the user to program his own file management system, by providing an appropriate environment in which the users program can operate. This would give the user maximum freedom in the handling of his on-line data base.

The direct application of computer techniques to biological research in collaboration with N.I.H. scientists led to the development of techniques for the storage, search and display of chemical structures encoded in the Wiswesser Line Notation. This work, in turn, led to the encoding of chemical structures as connection tables and a more user oriented chemical information system, for the search, storage and display of these data bases. A more complete utilization of data bases was the result of an interactive system that provided for an instantaneous retrospective search, of chemical literature, for occurrences of certain terms.

The program to educate scientists and administrators in the use of information science techniques and the Computer Center continues to grow. Over 1700 students attended the 80 courses offered in the Computer Training Program in the past year. Several hundred others could not be handled due to lack of sufficient staff.

Educational, career and advancement opportunities for NIH employees under the DHEW's New Careers Program was implemented in the form of a Computer Operator Training Program. Students were selected from over 100 candidates by the NIH merit promotion

plan, and will be trained in the theory and operation of large computer systems for one year. This provides these people with the opportunity to learn and develop to the maximum of their capability.

Serial No. 2.1
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: User Support and Communications

Previous Serial Number: Same

Principal Investigator: Frances E. Halverson

Other Investigators: Staff of PAL Unit

Cooperating Units: None

Man Years:

Total: 7
Professional: 7
Others: 0

Project Description:

Objectives:

To provide the users of the Computer Center with the personal assistance necessary if they are to make effective use of the Center's facilities. To provide users assistance in resolving problems encountered while using the systems (hardware and software) maintained by the Center.

Methods Employed:

The PAL (Programmer Assistance and Liaison) Unit was established to perform this function. Thru the Bugs, Diagnostics and HINTS SECTION OF THE Computer Center's technical report INTERFACE and the Users Guide, the PAL Unit keeps the user community abreast of the latest word in problems, system shanges and programming hints. They also assist the Systems Team in the design, selection installation, testing and modification of all system software provided by the Computer Center. They are in constant contact with IBM to insure that software problems are fixed adequately and quickly.

Significance to Program of the Division:

The computer user at NIH has at his disposal a group of competent professional programmers that are able to assist him with his problems in running programs or using the computer facility. The PAL Unit notes all trouble areas and, through INTERFACE, communicates common problem areas to all users. In all facets of computer use the user has someone that can give him the answer or guarantee to find the answer for him.

Honors and Awards: None

Publications: None

Serial No. 2.2
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Center Users Guide

Previous Serial Number: Same

Principal Investigator: Frances Halverson

Other Investigators: Leslie Barden, PAL Unit and Systems Team

Cooperating Units: None

Man Years:

Total: 1
Professional: 1
Others: 0

Project Description:

Objectives:

To provide the users of the central computer facility with a guide to the services, standards and use of the Computer Center.

Methods Employed:

A 200-page Users Guide was published and distributed to all users of the Computer Center, to Institute and Division Administrative or Executive Officers, to contract companies required to use the Computer Center and to other organizations and individuals having a logical need for it. Updates to the Users Guide are published and distributed as necessary to keep it current.

Significance to Program of the Division:

For the first time all information pertaining to computing was brought together in a single reference document. The Users Guide contains a description of all facilities and services and how to use them. Programming standards, languages supported, JCL summary and other facets of computing are all given in detail. It is a complete guide to computing for the computer user.

Honors and Awards: None

Publications: None

Serial No. 2.3
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: HASP

Previous Serial Number: 2.4

Principal Investigator: James Oberthaler

Other Investigators: Roger Fajman, William Jones

Cooperating Units: None

Man Years

Total: 1.2
Professional: 1.2
Others: 0

Project Description:

Background:

The HASP system is an extension to the principal control program employed on each of the Computer Center's IBM System/360 computers. HASP is responsible for reading all jobs into the system, scheduling their execution on a priority basis, and handling all printed and punched output from the jobs.

Objectives:

1. To tailor the standard HASP distributed by IBM to the specific needs of the NIH computing community.

2. To integrate the NIH modifications to Hasp into the new versions of HASP which are released periodically by IBM.

Methods Employed:

HASP is a computer program written in IBM System/360 Assembler Language. The standard version distributed by IBM consists of approximately 10000 instructions to which NIH has now added on the order of an additional 8000 instructions.

Major Findings:

A major modification completed during the past year will allow all of the center's computers to share a common work queue. This will effectively provide a single logical input work stream to all the computers. Jobs will be processed by the next computer ready to accept them, thus relieving the processing bottlenecks which occur when one processor is overloaded while others can accept additional work.

Significance to Program of the Division:

The multiple computer environment at the Computer Center allows significantly better overall service to the NIH computing community by providing both increased capability and superior reliability to a single machine operation. Within this framework, however, the task of managing three computers rather than one presents formidable scheduling and logistical problems. Through the use of a shared input/output queue, many of the decisions currently being made manually can be automated, thus providing more immediate and accurate response to the variations in work load and resource availability. In addition to the reduced turnaround time which should result, two other significant advantages will be apparent to users whose work locations are remote from the computer center.

1. Up to the present time, the computer center has required class C work (i.e., those jobs which are capable of tying up system resources for extended periods of time) to be submitted in person at the computer center to avoid the load balancing problems which could result if these jobs were submitted from Remote Job Entry (RJE) terminals or from WYLBUR/CPS terminals. This restriction may now be lifted.

2. It will also be possible for users to submit jobs through WYLBUR and have the output printed on any nearby RJE terminal.

Honors and Awards: None

Publications: None

Serial No. 2.4
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: 360 Systems Development

Previous Serial Number: 2.5

Principal Investigator: Robert H. Brunelle

Other Investigators: Staff of Systems Team

Cooperating Units: PAL Unit

Man Years:

Total: 6
Professional: 6
Others: 0

Project Description:

Objectives:

To maximize the thruput and minimize turnaround time to all users of the central facility's 360 computers. To put computing power into every researcher's lab or office thru remote terminals and software systems. In general, provide the NIH user with the best Computer Center and service he can get thru software development and hardware expansion.

Methods Employed:

Judicious selection and tailoring of software systems to the NIH environment. Acquisition and development of software and hardware to provide the researcher with the tools he needs. Constant attention to overall system software and hardware needs so that new systems and hardware are available when needed. Increasing the capacity of the system to keep ahead of the needs of the NIH users.

Significance to Program of the Division:

Constant attention to all facets of computing provides the NIH with an up-to-date computer facility catering to the needs of all NIH's researchers. Computer power in the lab brings a powerful research tool closer to the project, thus making it easier to use, and more likely to become an integral part of, the research program. All of these efforts combine to form a forward looking Computer Center sensitive to the needs of the NIH research environment.

Honors and Awards: None

Publications: None

Serial No. 2.5
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: WYLBUR Text-editing System

Previous Serial Number: 2.6

Principal Investigator: Roger Fajman

Other Investigators: Jennifer Fajman, Leslie Barden

Cooperating Units: None

Man Years:

Total: 1.5
Professional: 1.5
Others: 0

Project Description:

Objectives:

1. To provide Computer Center users with a convenient and comprehensive tool to aid in the creation and development of programs.

2. To make it significantly easier to create and edit text materials such as letters, reports, etc.

3. To reduce the necessity for users to leave their laboratories or offices in order to use the computer.

4. To provide a convenient method for ascertaining the status of the computing system as a whole and locating particular jobs as they are being processed.

Methods Employed:

WYLBUR is an on-line text-editing and remote job entry system oriented to low-speed character mode terminals such as the IBM 2741, 1050, and 2260 and Models 33, 35 and 37 teletype machines. It was originally designed

and implemented at the Stanford University Computation Center. WYLBUR provides the user, in his home or office, the facility to create and edit source programs in real time, submit them for compilation and execution by the standard job stream processor, and retrieve the results of execution at his terminal. In addition, the user may work with arbitrary text material, such as letters, reports, books, etc. WYLBUR is constantly being modified to improve performance and make new facilities available to the users.

WYLBUR consists of two computer programs written in IBM System/360 Assembler Language which reside permanently in two separate regions under OS/360. MILTEN, the telecommunications monitor, controls all input and output for the low-speed terminals and handles communication with the system operator. MILTEN can communicate with one or more subsystems operating in other regions under OS/360. all subsystems use the same pool of low-speed lines, thus reducing the cost and providing more flexibility for the user. WYLBUR, the text editor, is a subsystem of MILTEN. It performs all of the actual manipulation of text and the remote job entry to HASP. WYLBUR calls on MILTEN in a device independent manner for all terminal I/O requests. The permanent text files are stored on 2314 disks which are shared among all CPU's in the system. This permits batch jobs to access WYLBUR data sets. The text is stored in a compressed format in order to save space. The working files are stored on 2301 drum (with disk as backup) and are moved in and out of core under a demand paging algorithm.

Major Findings:

WYLBUR has been extensively modified in the last year in order to increase its usefulness to users and to increase its capacity.

1. The number of telephone lines into WYLBUR has been increased from 32 to 92 in order to keep pace with the constantly increasing demand. The maximum number of users simultaneously logged on is 86.

2. The number of online file disks has been increased from 16 to 22 in order to satisfy the increasing demand for storage space.

3. The FETCH, PRINT, and PURGE commands provide a significant increase in capability by permitting users to easily inspect job output at the terminal.

4. Extensions to the ALIGN command and the new JUSTIFY and CENTER commands greatly increase WYLBUR's usefulness for editing documents.

5. The CHANGE command was extended to provide additional facilities for numbering and the manipulation of tabular data.

6. Options on the RUN command and the new ROUTE together with the Shared Spool facilities of HASP, permit users to get the output from WYLBUR jobs on remote high-speed printers, thus eliminating the need for many users to ever come to the Computer Center.

7. The load-balancing features of the Shared Spool version of HASP permit long production jobs to be submitted through WYLBUR. This makes the preparation of such work much easier.

8. Options have been implemented which permit the user to exercise control over when, and in what order, his jobs are run.

9. The Model 37 teletype is supported at 10 characters per second.

10. Many extensions have been made to the procedures which can be used to perform useful functions for WYLBUR users in the batch job stream.

11. Other commands and options have been modified in order to increase their usefulness.

12. Multiple paging files permit WYLBUR to make use of the additional temporary storage space needed to support many more users.

In addition to the effort described above, there is an extensive program to educate the users about WYLBUR. Many sessions of the WYLBUR for Programmers and the WYLBUR for Secretaries and Administrative Personnel courses have been taught, new edition of the WYLBUR Manual has been published, and a WYLBUR Reference Card is nearly ready to be published.

Significance to Program of the Division:

WYLBUR represents a significant move towards lowering the amount of non-programming overhead involved in the development of programs. With the job output available

at the user's terminal, effective turnaround time is substantially lowered. In addition, the added ease of using a typewriter-like terminal instead of a keypunch contributes to more rapid development and debugging of programs. WYLBUR has also gained wide acceptance as a tool to aid in the preparation of memos, reports, and other documents. In addition, WYLBUR has been used for data collection and inquiry applications.

Honors and Awards: None

Publications: None

Serial No. 2.6
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: DATASTØR PROGRAM

Previous Serial Number: 2.8

Principal Investigator: James Oberthaler

Other Investigators: William Jones

Cooperating Units: None

Man Years

Total: 1/2
Professional: 1/2
Others: 0

Project Description:

Objectives:

1. To provide specifications for a general purpose processor-to-processor communications facility.

2. To implement, using these specifications, a set of computer programs which will permit computers remote from the main CCB facility to create, manipulate, and retrieve data files on the computer center's disk storage devices.

Methods Employed:

1. Communications links between the central and remote sites may utilize switched network or leased line facilities; both synchronous and asynchronous transmission techniques are supported.

2. The programs which compose the DATASTOR system are coded in System/360 assembler language.

3. The programs operate in the batch environment at present, as opposed to being a permanently on-line system, such as WYLBUR.

Major Findings:

1. A working version of the specifications for processor-to-processor communication has been developed, and is available from the Computer Center. The title of this specification is DATASTOR and the Lavender Box.

2. A preliminary version of the system has been installed and is currently being used by the Division of Research Resources' AGT-30 computer. This version incorporates a significant and functionally complete subset of the facilities proposed in the above mentioned specification. The communications link with the AGT-30 operates at 150 characters per second over the switched network and utilizes the Bi-synchronous transmission technique.

3. Tests have been performed to verify the feasibility of connecting the Clinical Center's Programmed Console (PC) computer to the DATASTOR system. This link would utilize asynchronous data transmission via a leased line.

Significance to Program of the Division:

This is an important step in the Computer Center's plan to create an environment in which the many small specialized laboratory computers can easily converse with the powerful central facility.

Proposed Course:

1. To complete the implementation of the system according to the specfcation developed.

2. To develop a facility within the framework of the system for remote computers to submit batch jobs and receive their output.

3. To institute permanently on-line service by the system, as soon as its utilization is sufficient to justify doing so.

Honors and Awards: None

Publications: None

Serial No. 2.7
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Pharmacology Project, AGT-30

Previous Serial Number: 2.9

Principal Investigator: Charles R. T. Bacon

Other Investigators: J. Kaufmann, R.W. Moseley

Cooperating Units: DRR

Man Years

Total: 1-1/2
Professional: 1-1/2
Others: 0

Project Description:

Objectives:

The system is being developed collaboratively by DRR and DCRT as a research tool for pharmacology and toxicology. The project is a first step in DRR's development and evaluation of an integrated set of computer techniques for handling data on chemical compounds and the functioning of living systems. The DCRT effort is more generalized, so that resulting computer systems may be useful to a wide community of NIH computer users.

Methods Employed:

The AGT-30 display terminal consists of a computer with a CRT, designed to display three-dimensional data with maximum realism. It displays a two-dimensional projection of three-dimensional data stored in its memory, and the operator may, by use of control dials and pushbuttons, alter viewing angles, magnifications, relationships of picture parts to one another, etc.

Pictures may be created by programs running on the IBM 360 computers, stored as datasets, and transmitted to the AGT-30 display terminal over telephone lines. Interaction between the display operator and the program generating the picture is not possible, but is envisioned for future implementation. A paper describing the 360 picture language is available from the Technical Information Office, DCRT, Building 12, Room 2235.

In addition to the picture communication and display programs written for the AGT-30, a system of programs has been written to permit chemists to sketch molecular diagrams in two dimensions, labeling atoms, permitting double, triple, steric, and resonant bonds to be depicted, and resulting diagrams to be stored locally on a disk file. The underlying structures are explicitly maintained by the sketching program, such that programs may be written which analyze these structures, with direct reference to atom and bond tables. In addition, there is provision for inclusion of numeric information about the chemical properties of the molecule.

Both major systems are available to users presently, but a most desirable link, permitting 360 programmers access to the molecular-structure data, is still incomplete. Further work in both systems is continuing, although they are presently usable.

Significance to Program of the Division:

The possibilities of three-dimensional picture display are limited by the systems available to potential users. The link between the 360 programmer and the AGT-30 has been completed, enabling a wide range of graphic explorations to be undertaken.

Input of chemical structures is a topic of great interest, also. At the time Bill Moseley undertook to write a program for the AGT-30 for this purpose, that feat seemed quite large. Now, almost a year later, Feldmann's work on the PDP-10 (Annual Report Nos. 2.27 and 2.30) uses a similar technique as part of a substructure search program. It is not appropriate to attempt chemical structure analysis or search on the AGT-30, and hence our aim is to provide a structure format and data link which will be accessible to 360 programmers.

Honors and Awards: None

Publications:

1. Bacon, C.R.T.: 3DP: Three-Dimensional Pictures - "A 360 Graphic Language for AGT-30 Display," Computer Center Branch Internal Memo; March, 1971.

Serial No. 2.8
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: INTERFACE (technical newsletter)

Previous Serial Number: 2.12

Principal Investigator: Joseph D. Naughton

Other Investigators: Staff of Systems Team and PAL Unit

Cooperating Units: None

Man Years:

Total: 2
Professional: 1
Others: 1

Project Description:

Objectives:

1. The primary purpose of this effort is to give the Computer Center a regular means of conveying, (1) highly current technical information to the widely dispersed computer programmers and systems analysts at NIH, and (2) current Center accomplishments, plans, policies, course offerings and other information of interest to personnel and general managers at NIH.

2. A secondary purpose is to foster communication and collaboration among computer users and between them and the Center by providing a forum for: (1) announcing key new projects, accomplishments, personnel or organizations; and (2) airing viewpoints or suggested approaches to computing problems.

Methods Employed:

INTERFACE is published every three-to-four weeks, and distributed to all scientific and administrative personnel

who have expressed a desire to be kept up-to-date on computing at NIH.

Significance to Program of the Division:

INTERFACE complements the Users Guide and other technical manuals by highlighting items and directing readers to the other publications for details. INTERFACE has given the users of the Computer Center a single reference point for all communications concerning the use of computers at NIH. It has done an excellent job of keeping the users informed of new services, major systems changes and all facets of computing at NIH.

Proposed Course:

INTERFACE will continue to inform the Computer Center users of all pertinent computing highlights on a regular basis.

Honors and Awards: None

Publications: None

Serial No. 2.9
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: IBM 2250 Graphic Program Support

Previous Serial Number: 2.13

Principal Investigator: James Oberthaler

Other Investigators: Jennifer Smith, Robert H. Brunelle

Cooperating Units: None

Man Years

Total: 1/4
Professional: 1/4
Others: 0

Project Description:

Objectives:

1. To maintain the NIH-produced PL1 Graphic Subroutine Package which enables users to communicate with the IBM 2250 Graphic Display using the PL1 programming language.

2. To instruct potential users in the Graphics capabilities offered by the Computer Center.

3. To act as a consultant for programmers implementing systems for the 2250 Display.

4. To study the graphic computing needs of NIH and implement software systems to fulfill these needs.

Methods Employed:

The PL/1 Graphic Subroutine Package is designed to enable the programmer to communicate with the IBM 2250-1 using

OS PL/1, version 5. The programmer using this package can have access to the 2250-1, as well as the resources of PL/1. The ability to issue calls to procedures written in assembly language enables the programmer to send data to and from the display. Interrupts are handled by user-defined PL/1 procedures. The subroutines included in the IBM System/360 Operating System Graphic Programming Services for the IBM 2250 Display Unit, Form C27-6909, are also made accessible to the programmer.

Major Findings:

1. The teaching of Fall and Spring semester courses in:
<u>Basic Graphics</u>, and,
<u>PL1 Graphics</u>.

2. The changing of the Graphics environment from a permanently on-line system to a batch access system:

a. made it feasible to execute the program in high-speed memory reducing computing time by approximately two-thirds.

b. made it possible for the user to access arbitrary disk and tape data sets.

c. allowed the user to select the amount of main storage necessary to perform his computations.

Significance to Program of the Division:

NIH users will be able to extend the data processing power of System/360 computers: (1) to handle the graphic information associated with medical research and analysis applications; and, (2) to provide faster and more effective retrieval and graphic expression of medical data.

Proposed Course:

1. To implement Graphics support for the Fortran Language.

2. To connect the 2250 Display to one of the Computer Center's IBM System/360 Model 65's in order to provide better response time for the user.

Honors and Awards: None

Publications: None

Serial No. 2.10
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: CPS - Conversational Programming System

Previous Serial Number: 2.14

Principal Investigator: John Camp

Other Investigators: Trudy Kenny

Cooperating Units: None

Man Years:

Total: 1-1/2
Professional: 1-1/2
Others: 0

Project Description:

Objectives:

CPS continues to provide much of the power of the central computer facility to the researcher in his lab or office. By using CPS the researcher has available the computing capability of the central 360 computers as well as the ability to access data bases stored on-line at the central site.

Methods Employed:

The CPS system allows multiple users to simultaneously write, debug, and execute 360 computer programs conversationally. Programs are written in a subset of the PL/1 language from a 2741 typewriter terminal or teletype located in or near the user's office. The terminals are are connected with the NIH central 360 system over telephone lines on a dial-up basis. In addition, CPS supports a dialect of the BASIC language and provides a remote job entry (RJE) facility through which jobs can be submitted to be run in the background batch environment.

Major Findings:

Continuing enhancements to CPS place increasing power at the user's disposal. During the past year three versions of CPS have been installed. Each of these has increased the reliability of the system and added additional power to the language. Examples of new language features are: edit directed input/output, support of controlled storage, and a method of simulating terminal input and output. It is now possible to write programs which dynamically modify themselves during execution. A series of built-in functions have been added which allow a user to process records containing mixed data types.

NIH modifications to this IBM distributed program have enabled it to operate effectively as an integral part of the Computer Center. These modifications provide interfaces for accounting, remote job entry, and dataset naming.

During the past year there has been continuing growth in the number of users and in the use of CPS. Availability of the system to users was extended to the full normal operating day of 8:30 a.m. to midnight. The numbers of active users of CPS has increased by one third over the number a year ago.

Significance to Program of the Division:

CPS continues to provide computing capacity in the lab or office allowing the researcher to utilize the central computer facility without leaving his normal working environment. Thus the computer is more readily available and much time is saved by not making the user come to the central site. By providing a truly interactive conversational environment CPS gives the user capabilities which could never be achieved in a strict batch mode of operation.

Proposed Course:

Intended support includes continuing to provide maintenance for CPS. New versions and updates will be installed as they become available and are needed to improve the reliability and performance of the system. Local plans include steps to further integrate CPS into the operation of the central facility, improvement in user interaction with CPS, and improved performance.

Honors and Awards: None

Publications: None

Serial No. 2.11
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: PDP-10 Systems

Previous Serial Number: 2.16

Principal Investigator: H.W. Vreenegoor

Other Investigators: S. Bryan, R. Grunby

Cooperating Units: CSL, HL

Man Years:

Total: 1-1/2
Professional: 1-1/2
Others: 0

Project Description:

Objectives:

1. Development and improvement of new and existing PDP-10 monitor features and hardware functions.

2. Development and implementation of techniques for improving computer-computer communications.

3. Development of remote graphics terminals as a tool for the biomedical investigator.

Methods Employed:

Continuing development of the time-sharing monitor aimed at increasing its reliability and response to the NIH operating environment and job mix. Also, the program library was augmented by several important system programs such as a powerful text editor (SOS) and an Algol like compiler (SAIL).

The equipment configuration was augmented by: (1) A module of 1.8 microsecond, 128K words of mass memory and; (2) 6 dispack drives with a capacity of 5.5 million words/pack and a transfer time of 15 microseconds/word.

Hardware has been designed to permit the construction of a PDP-10/IBM360 interface. This interface will permit the PDP-10 computer to access the IBM360 via the datastore module for file storage and retrieval.

A remote graphics device is on order and is expected to arrive in the late spring of 1971.

Major Findings:

The PDP-10 monitor has continued to prove itself as a reliable and dependable system. The amount of downtime due to a faulty functioning of this monitor has been negligable. Also, downtime due to major equipment malfunction has been extremely low.

The addition of diskpack storage has greatly increased the usefulness of the system to remote users since programs and data can now be stored on-line.

Significance:

At this time, 5 laboratory-based computers communicate with the PDP-10 computer over normal telephone lines. It has been shown that the development of computer programs for laboratory-based computers is greatly facilitated by using the PDP-10 computer for the initial phases of this development. Also, the PDP-10 is used for file storage and computational backup which greatly enhances the capabilities of the laboratory-based systems.

The PDP-10 graphics capability is becoming increasingly popular and is used by several NIH laboratories involved in molecular research and cancer cell growth. Some of this work could not have been started without the capabilities available on this interactive display system.

Proposed Course:

Continuing development and exploitation of the unique graphics and computing facilities offered by the time-shared PDP-10 system with emphasis on the development of higher speed communications between laboratory computers and DCRT computers and the development and implementation of software for remote graphics devices.

Honors and Awards: None

Publications: None

Serial No. 2.12
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Training Program

Previous Serial Number: None

Principal Investigator: Rita G. Minker

Other Investigators: Patricia L. Logan

Cooperating Units: All DCRT Laboratories and Branches

Man Years:

Total: 6.75
Professional: 6
Others: .75

Project Description:

Objectives:

1. To provide the NIH staff with an understanding of and ability to use modern computer technology.

2. To facilitate the formulation and analysis of the problems to be solved so that the technology can be usefully applied.

Methods Employed:

1. Training courses are tailored to the specific hardware and software available at NIH.

2. Seminars focus on both the underlying disciplines (e.g., applied mathematics, computer-related engineering, information sciences) and on the specific areas of potential application.

Significance to Program of the Division:

Training courses enable investigators to use DCRT's powerful equipment. Effective use of advanced techniques

and advanced equipment is obtained by further training and directed discussion.

During the past 12 months, over 1700 registrations were processed for the approximately 80 short courses offered.

Proposed Course:

Fall 1971 and Spring 1972 semesters of training classes and seminars are planned.

Honors and Awards: None

Publications:

1. Computer Center Branch: Computer Training Courses--Fall Term 1970. U. S. Dept. of Health, Education and Welfare, Public Health Service, National Institutes of Health, Division of Computer Research and Technology, Computer Center Branch.

2. Computer Center Branch: Computer Training Courses and Seminars--Spring Term 1971. U. S. Dept. of Health, Education and Welfare, Public Health Service, National Institutes of Health, Division of Computer Research and Technology, Computer Center Branch.

3. Computer Center Branch: Introduction to Services at CCB.

4. Computer Center Branch: Teach Yourself CPS-A Self Study Guide.

Serial No. 2.13
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: On-line File Processing System

Previous Serial Number: None

Principal Investigator: Jon A. Halverson

Other Investigators: Marvin Katz, Richard Baxter

Cooperating Units: None

Man Years:

Total: 2-1/2
Professional: 2-1/2
Others: 0

Project Description:

Objectives:

The primary objective is to establish an on-line file processing capability at the NIH whereby the user can maintain, retrieve and display information in a data base via a remote terminal.

Methods Employed:

The system being implemented is a modularly constructed program designed to provide the basic control system program for the installation of on-line information systems. It functions as an interface between user-written processing programs and the IBM system/360 operating system.

Major Findings:

The basic system has been implemented and several sample user programs, written in COBOL, PL/1 and Assembler have been run in simulation, using the card reader and printer.

Significance to Program of the Division:

The system can be considered as an initial step toward the development of on-line information systems for the NIH. For the first time investigators responsible for maintaining large data bases will have a powerful means of accessing critical information quickly.

Proposed Course:

Attention will be given to operating the system in an on-line, multi-user environment. Special emphasis will be given to applying the system to those areas in the NIH Clinical Center which can benefit from interactive processing.

Honors and Awards: None

Publications: None

Serial No. 2.14
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Optical Document Processing System

Previous Serial Number: None

Principal Investigator: Elliott Alterman

Other Investigators: John Camp

Cooperating Units: None

Man Years:

Total:	.9
Professional:	.9
Other:	0

Project Description:

Objectives:

NIH processes vast volumes of data in many forms and formats. The effort necessary to analyze and process the data currently requires significant time and effort to convert these data to machine readable form. We have eliminated much of this intermediate data conversion by using the documents themselves as input to the central facility computers.

Methods Employed:

Internal processing routines have been developed to process documents read by the optical character recognition equipment (IBM 1288 Optical Page Reader). Each document is described uniquely in terms of a set of parameters. This information will be maintained in an on-line data set accessible to the processing program. The flexibility of this design is displayed in the elminiation of the requirement for writing new processing programs for each new document we wish to process. Changes to document design, processing options, or output formatting can be done in a matter of minutes.

Significance to Program of the Division:

This system permits rapid and efficient collection of data for processing by a computer. Data usefulness is thus improved by timely availability.

Honors and Awards: None

Publications: None

Serial No. 2.15
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Search, Storage and Display of Connection Table Data Bases

Previous Serial Number: None

Principal Investigator: R.J. Feldmann

Other Investigators: G.F. Hazard, S.R. Heller

Cooperating Units: DMB, PRA-NCI

Man Years:

Total: 1.3
Professional: 1.3
Others: 0

Project Description:

Objectives:

The development of techniques for storage, search and display of chemical structures encoded as connection tables. The planning of a user oriented chemical information system.

Methods Employed:

Programs were written and debugged for:

1. Processing of connection table encoded files
2. Sub-structure search of processed connection table files
3. Display of chemical structures

Consideration was given to the types of interaction involved in specifying sub-structure search queries.

Major Findings:

The results obtained from searches of connection table data bases are superior to equivalent searches of data bases encoded in wiswesser line notation (WLN). Similarly the connection table of a compound provides a superior base from which to generate a two dimension display or printed diagram.

Significance:

The interactive style in which the programs are written permit a user to answer in seconds to minutes queries which formerly took from days to weeks to accomplish by computer.

Proposed Course:

The existing programs will be improved to permit more accurate and relevant searches to be performed. Work will be done to integrate the programs and files to create a working chemical information system.

Honors and Awards: None

Publications: None

Serial No. 2.16
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Retrospective Search of Chemical Literature

Previous Serial Number: None

Principal Investigator: R.J. Feldmann

Other Investigators: K.P. Shapiro

Cooperating Units: DMB

Man Years:

Total: 0.8
Professional: 0.8
Others: 0

Project Description:

Objectives:

The development of an instantaneous retrospective search capability of the CBAC (chemical biological activity) journal published by chemical abstracts service.

Methods Employed:

Scatter storage techniques (hashing) were investigated and applied programs were written and debugged for:

1. Processing CBAC tapes
2. Generating a HASH file and
3. Retrieving CBAC references from the HASH file

Major Findings:

A retrospective search of CBAC for the occurrences of a term (compounds, registry numbers, authors, etc) can be performed within 1 second. The interactive analysis of search results by the user leads to a more

complete utilization of the data base and the retrieval results. The present implementation obtains adequate coverage of this abstract journal by storing only the reference (volume, issue, abstract number) to the CBAC abstract.

Significance:

When all the sections of chemical abstracts are produced by computer the CBAC format will be used. The experience gained with CBAC will lead to a more complete computer coverage of chemical literature.

Honors and Awards: None

Publications: None

Serial No. 217
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Dial-Up Graphics

Previous Serial Number: None

Principal Investigator: R.J. Feldmann

Other Investigators: None

Cooperating Units: None

Man Years:

Total:	0.2
Professional:	0.2
Others:	0

Project Description:

Objectives:

The development of a computer terminal with Graphics capability. Communication with the computer (PDP-10) to be over dial-up telephone lines.

Methods Employed:

An analysis of the characteristics of existing display and displays in development indicates must have a tablet (Rand Tablet) to be interactive. Consideration has been given to the changes to the PDP-10 system software needed to support remote (Dial-up) graphics units.

Major Findings:

The developing state of technology has produced several technologically different types of graphics terminals. The silicon target storage tube technology makes possible a graphics terminal which is highly interactive (rapidly changing flicker-free bright images).

Significance:

The dial-up graphics terminal when purchased and implemented will provide the input-output mechanism for chemical structure diagrams, the reproducibility of the dial-up graphics terminal will enable the chemical information system to be accessed by scientists at the site of their work rather than at a computer center.

Honors and Awards: None

Publications: None

Serial No. 2. 18
1. Computer Center Branch
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Wiswesser Line Notation (WLN) Search and Retrieval

Previous Serial Number: None

Principal Investigator: R. J. Feldmann

Other Investigators: D. A. Koniver

Cooperating Units: HL

Man Years:

Total: 0.4
Professional: 0.4
Others: 0

Project Description:

Objectives:

The development of techniques for storage, search and display of chemical structures encoded in the Wiswesser Line Notation (WLN).

Methods Employed:

Programs were written and debugged for:

1. Processing WLN Encoded Files
2. Searching a processed WLN file to obtain compounds with specified fragments
3. Display of structures coded in WLN

Major Findings:

The WLN is used quite extensively in industry because it requires the least computer equipment (if any) to implement and search a file. The WLN for biologically

interesting compounds is quite complex. By contrast the connection table approach used by chemical abstracts service requires no adjustment for large or complex compounds. Because of the way the WLN represents the connectivity of a compound certain types of searches are very difficult to state and perform.

Significance:

As the chemical information system begins to develop an adequate internal computer representation must be chosen. The experience gained from this project led to a parallel investigation of connection table data bases.

Proposed Course:

The programs as developed will be maintained and will be available to any user who has WLN encoded files.

Honors and Awards: None

Publications:

1. R. J. Feldmann, S. R. Heller, and K. P. Shapiro: 5th Computer Aided Experimental Spectroscopy Study Group, Midland, Michigan; November 4-6, 1970.

2. S. R. Heller: invited speaker, Chemical Abstracts Service Advisory Board Meeting, Columbus, Ohio; November 5, 1970.

3. R. J. Feldmann, S. R. Heller, K. P. Shapiro, and R. S. Heller: Abstracts of Papers, CHLT 21, 161 ST. ACS National Meeting, Los Angeles, California; March 28-April 2, 1971.

4. R. J. Feldmann: invited speaker, World Health Organization Workshop on an International Reference Center Network for Psychotropic Drugs, Plitvice, Yugoslavia; June 21-25, 1971.

5. S. R. Heller: invited speaker, World Health Organization Workshop on an International Reference Center Network for Psychotropic Drugs, Plitvice, Yugoslavia; June 21-25, 1971.

July 1, 1970 through June 30, 1971

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT - 3

2. LABORATORY OF APPLIED STUDIES — 3. Eugene K. Harris, Chief

I. GENERAL SUMMARY

The Laboratory of Applied Studies engages primarily in cooperative studies, with NIH scientists, applying mathematics, statistics, and computing science to biomedical problems. To support these studies, special-purpose computer programming systems are developed where needed; these are often later extended for more general use and distributed to potential users in the form of DCRT technical reports. In addition, the Laboratory undertakes mathematical and mathematical statistical research with applications to biological studies.

Continuing last year's format, each project report includes a number of "subprojects" contributing to the same general objective. One new project, begun this year, involves the use of computer simulation techniques and hardware models in neurophysiological research.

II. CURRENT LABORATORY PROGRAMS

1. Clinical Projects

As a result of several years of collaborative research efforts by LAS with Institute or Clinical Center scientists, various projects have now matured to the stage of direct application to patient care. Chief among these are the cardiology projects discussed in Project Nos. 3.2 and 3.3. Specifically, these include the use of computer programs to aid in evaluating ECG's and the application of specially designed data acquisition devices, hemodynamic programs, and mathematical methods to assess the functional status of the heart.

In the latter project, computer processing and video tracking are combined to achieve continuous measurements of ventricular volumes and intraventricular pressure. Replacement of manual planimetry by these automated procedures represents improved precision and enormous savings of professional time. Moreover, such vital cardiac parameters as stroke volume, work, ventricular contractility, etc., can now be estimated with much greater ease and accuracy from continuous records.

Computer-based ECG diagnosis is not so clearly a distinct improvement over manual review and analysis. An ongoing LAS activity, in cooperation with the Cardiology Branch, NHLI, is the implementation and comparative study of major computer systems for automated analysis of ECG's from the standpoints of sensitivity, consistency, and cost-effectiveness as an adjunct to manual diagnosis. Two such systems are already in routine use at NIH (the Mayo Clinic program and the

Public Health Service program); a third will be brought up during FY 72. The best system may require a synthesis of existing features.

Statistical research of analytic and biological components of variance in blood constituents of normal individuals (Project 3.4), conducted during the past several years by LAS in collaboration with the Clinical Pathology Department, CC, is now also turning towards patient care applications. The specific concern at this time is the evaluation of drug therapy for hypertension through the analysis of blood pressure and serum electrolytes sampled on a round-the-clock basis. The collaborating clinical laboratory is the Endocrinology Branch, NHLI. Statistical methods will emphasize the study of rhythmicity in both clinical status and sensitivity to therapy. More general statistical research will use the existing multivariate data collected from normal volunteers to study the significance of inter-individual variation in setting normal ranges for diagnostic purposes.

A new collaborative study in the application of computer technology to patient care involves a three-way collaboration between LAS staff, the Nuclear Medicine Department of the Clinical Center and Institute clinicians. A combined gamma camera-computing system secured during this reporting year will be used to study fluid flow, transport and diffusion processes following administration of small amounts of the radioisotope Technetium 99m. The computer's extremely rapid data manipulation and storage facilities permit much finer image definition for diagnostic purposes. For example, non-linearities in the relationship between the number of scintillation counts and the intensity of image are no longer a problem when a computer is used. Further, the computer will facilitate dynamic studies of vascular flow patterns in both normal and pathologic conditions, particularly in response to therapy.

The essential LAS contribution to these clinical studies is the multidisciplinary team approach. The LAS group includes a board-qualified internist, electronic engineers, mathematicians and statisticians, computer specialists and programmers, drawn from all sections and units of the Laboratory and other DCRT labs and branches. Individual special talents are pooled to gain a greater understanding of the biomedical problem as well as to test and implement the necessary computing methods. The association with the NIH scientist(s) then becomes a true cooperation in the performance of research, not simply a consulting or service function.

2. Laboratory Research Projects

A highly productive LAS effort this year in support of laboratory research at NIH has been the contribution of the Applied Mathematics Unit to the mathematical theory of macromolecule-ligand binding, in particular fatty acid-protein binding. (Project No. 3.5).

The well-known Scatchard binding model (a weighted sum of binding constants) has been shown to be mathematically equivalent to the step-wise series of equilibrium reactions corresponding to individual binding sites. A computer algorithm has been developed, using least squares estimates of the parameters of the Scatchard model, to determine the dissociation coefficients of the

step-wise equations without having to assume independence of binding sites or other restrictive conditions. It now becomes worthwhile to reanalyze published data previously fitted by the Scatchard model alone. This reanalysis is underway and should lead to valuable new information concerning the dissociation coefficients of many macromolecule-ligand reactions of physiologic importance.

The LAS collaborative program in neurophysiology has been concentrated this year in two projects where computer simulation and modelling have proven useful for experimental investigations. The first of these concerns the development of a hardware neural model which reproduces the natural response to a distributed input (analogous to a dendritic net). The model can simulate both spatial and temporal summation of input as well as the electrotonic spread of an action potential or postsynaptic potential. These hardware devices can be linked together to provide a realistic basis for studying the characteristics of small neural nets.

The cerebellar cortex forms the subject of the second project. Hypotheses of interaction between mossy and climbing fibers, Purkinje cells and Purkinje recurrent collaterals, have been developed from a computer-simulated model of mammalian cerebellar cortex. These are being tested through analysis of electrophysiological records from the cerebellar cortex of monkeys trained to perform selected motor tasks. These behavioral studies, which will, in turn, lead to further experiments with the computer model, are being undertaken in cooperation with the Laboratory of Neurophysiology, NIMH. Hopefully, they will lead to a better understanding of the functional (in contrast to anatomical) organization of the cerebellum, largely unknown at this time.

3. Research Program in Mathematics and Statistics

The Biomathematics and Statistics Section of the Laboratory has continued its development of the theoretical foundations for analysis of multivariate ratios and proportions, designated "size and shape" theory (Project No. 3.1). The past year has seen a growing number of applications of this theory to biomedical research: for example, in the analysis of genetic factors related to cleft palate. The Section has renewed its work on the applications of graph and network theory to molecular biology, in collaboration with the Laboratory of Molecular Biology, NCI.

4. Education and Training

During the past year, Laboratory staff conducted courses for NIH scientists in the areas of time series analysis, differential equations, complex variables and the use of the CalComp plotter.

LAS reports published or "in press" during FY 71

1. Buchsbaum, M., and Harris, E. K.: Diurnal variation in serum and urine electrolytes. J. of Applied Physiology. 30: 1 27-35, January, 1971.

2. *Cotlove, E., Harris, E. K., and Williams, G. Z.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, III. Physiological and medical implications. Clinical Chemistry. 16: 12 1028-1032, 1970.

3. Fletcher, J. E., Spector, A. A., and Ashbrook, J. D.: Analysis of macromolecule-ligand binding by determination of stepwise equilibrium constants. Biochemistry. 9: 4580-4587, 1970.

4. *Harris, E. K.: Distinguishing physiologic variation from analytic variation. J. of Chronic Diseases. 23: 469-480, 1970.

5. *Harris, E. K., Kanofsky, P., Shakarji, G., and Cotlove, E.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, II. Estimating biological components of variation. Clinical Chemistry. 16: 12 1022-1027, 1970.

6. Hill, E.: The on line modeling system, Parts I, II, III. DCRT technical report No. 6.

7. Hutchinson, G.: Modular lattices and abelian categories. J. of Algebra. (in press)

8. Magar, M. E., Steiner, R. F., and Fletcher, J. E.: Analysis of protein ligand equilibria. J. Theoretical Biology. (in press)

9. Marcus, M. L., Schuette, W., Whitehouse, W., Bailey, J. J., Glancy, D. L., and Epstein, S.: A completely automated video-tracking technique for the determination of dynamic changes in ventricular volume. Circulation. (in press) (abstract: XLII: Suppl. III, 101, Oct., 1970)

10. *Minicozzi, W., and Stroot, M.: On the determination of interaction energy functions, II. Crystalline formic acid. J. of Computational Physics. August, 1970.

11. *Mosimann, J. E.: Discrete distribution models arising in pollen studies. In Random counts in physical science, geoscience, and business, Pennsylvania State University Press. 1-30 1970.

12. *Mosimann, J. E.: Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions, J. of the American Statistical Association. 65: 930-945 1970.

*paper cited in last year's annual report as being "in press"

13. Young, D. S., Harris, E. K., and Cotlove, E.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, IV. Results of a study designed to eliminate long-term analytic deviations, Clinical Chemistry. 17:5, 403-410, 1971.

Serial No. DCRT 3.1
1. Laboratory of Applied Studies
2. Biomathematics and Statistics Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Biomathematics and Statistics

Previous Serial Number: Same

Principal Investigator: J. E. Mosimann
Assisted by: G. Hutchinson, G. Atta

Co-Investigators: J. Niswander, HG, NIDR
P. Berthet, Univ. of Louvain, Belgium
P. Jolicoeur, Univ. of Montreal, Canada
A. Squillace, USDA

Man Years: Professional - 3.0
Other - 0.2

Objectives:

1) To conduct mathematical statistical research in areas of biological importance, particularly in a class of multivariate problems designated as "size and shape" problems.
2) To develop computing methodology of value to biostatisticians and other health scientists, particularly programs permitting the efficient use of multivariate size and shape methods.
3) To investigate the application of mathematical theory, particularly linear algebra and graph theory, to molecular biology and biochemistry.

Progress During Past Year:

a) Multivariate Size and Shape Statistics:

The goal of this continuing project is to provide a solid foundation in statistical theory for the study and interpretation of data which take the form of proportions, ratios, and indices of various form.

Size and shape variables are defined in a probabilistic context. Loosely speaking, dimensioned random variables like length, width, weight are "size" variables, while non-dimensioned variables such as the ratios height/length, or brain weight/body weight, are shape variables.

The most suitable problem for illustration of the concepts is that of size-related shape changes in man. The designation "size and shape problems" stems from this application. However, the methods are of general applicability in any field where ratios or proportions are useful. During the past year the methods have been applied to a variety of data sets: distribution of microarthropods in soil; human skull measurements to predict the probability of parents having children with cleft palate; chemical composition of monoterpenes to elucidate genetic mechanisms (see below); biomechanics of the hindlimb of the laboratory frog. In conjunction with these data sets, programs have been and are being, developed to make readily available the application of "size and shape" theory. Each of these analyses, from different fields, is being used to test in depth the theory and utility of size and shape methods.

In examining the inheritance of cleft palate, human skull proportions have been considered. By studying directly the relations of various facial proportions with simply defined size variables, it is possible to see that shape is related to size in a similar manner in the various experimental groups. Although there are shape differences between controls, and parents of affected children, these differences are apparently related to more basic underlying size differences. The relations of size to shape are similar for controls and experimentals, males and females. However, due to size differences among these categories, shape differences associated with size also occur.

Monoterpenes are simple organic molecules, directly related to the action of single genes. The analysis depends on "shape" variables (i.e., proportions, as in all chemical composition data) alone. Application of shape methods shows that simple genetic effects are readily apparent from the graphic output of the size and shape computer programs. Here, the biochemical problem of interest is twofold: first, the elucidation of certain dominant-recessive genes; second, an attempt to model possible synthetic pathways for the organic compounds involved, with hypotheses as to whether the action of various genes block or enhance reactions in certain directions.

From the point of view of theory, a number of new developments have been made in the past year, particularly in the area of the multivariate lognormal distribution, and generalizations of Dirichlet-generating distributions. The concept of neutrality (Connor and Mosimann, 1969) has been generalized and its dependence on a size-function made clear. A striking result states that at most one kind of "size-neutrality" can obtain for any given positive random vector variable.

The theoretical paper cited in last year's report as being in press, appeared, and is included again with this year's publications. Other papers, including one theoretical paper and empirical papers on at least three of the four data sets noted, are in progress.

b) Mathematical and Computing Methods in Molecular Biology and Bio-Chemistry:

Biomathematical applications of algebra and graph theory were studied. A major theoretical contribution, relevant to linear algebra, has been accepted for publication (Hutchinson, 1970). In addition, a follow-up study on related questions of computability has been completed.

A small project to test consistency of hypotheses with respect to deletion mutants of phage lambda was completed. A computer program tested more than one hundred fifty different hypotheses, showing which of these hypotheses gave the best explanation of recombination percentages obtained by experiment. The programming methods involved, using string manipulations, relate closely to algebraic areas of interest.

Recent work has centered upon the application of network theory and linear algebra to chemical or biochemical reaction systems. The set of reaction equations is often called the "mechanism" of the system. Certain conclusions about the behavior of a system can be obtained from the study of its mechanism. Especially in biochemistry, systems with very complex mechanisms have received more and more attention. An algebraic approach to chemical reaction mechanisms is under development by Hutchinson. It differs from standard theoretical approaches to reaction systems in that assumptions concerning temperatures, reaction rates, and flow of energy into or out of the system need not be made. Specifically, it is not assumed that the system goes to equilibrium. By arguments using linear inequalities it can be determined from a given mechanism that certain reactions must terminate, while other reactions may persist indefinitely. Linear estimates of the number of reaction occurrences before termination are obtained. Other results relate to the exhaustion of substances in the reaction system. It is intended to develop and adapt linear programming and graph theoretical methods for analysis of mechanisms. By such techniques, chemists and biochemists would have tools to obtain useful preliminary information concerning known or proposed new mechanisms.

Publications

1. Mosimann, J. E.: Discrete distribution models arising in pollen studies. In Random Counts in Physical Science, Pennsylvania State University Press 1-30 1970.

2. Mosimann, J. E.: Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions, J. of the American Statistical Association. 65:930-945 1970.

3. Hutchinson, G.: Modular lattices and abelian categories. J. of Algebra. (in press)

Serial No. DCRT 3.2
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Evaluation of Computer-assisted Systems for Patient Care

Previous Serial Number: Same

Principal Investigators: J. J. Bailey, M. Horton
Assisted by: M. Douglas

Co-Investigators: S. Epstein, D. Beiser, S. Itscoitz, CB, NHLI
G. Johnson, S. Larson, Nuclear Medicine, CC
D. Syed, H. Ostrow, CSL, DCRT

Man Years: Professional - 2.0
Other - 0.3

Project Descriptions:

Three sub-projects are currently being conducted in the general area of the evaluation of health care systems, particularly those concerned with cardiovascular problems. These projects are: a) evaluation of computer diagnosis of electrocardiograms (ECG); b) cost-effectiveness study of coronary care units; c) computer-assisted studies in nuclear medicine. The first two of these projects were introduced in last year's report.

a) <u>Automated ECG Analysis</u>

Background:

Recent studies by the developers of computer programs for ECG analysis have suggested that EKG reading time may be reduced 5- to 10-fold with consequent dollar savings in cardiologist's time. In addition, upgrading of EKG readings may result since the computer system makes all recommended measurements, is tireless and may include analytic statements neglected by the human. However, cost-effectiveness may vary considerably, depending on whether the population studied is largely normal or contains a high percentage of abnormal and bizarre cases. To evaluate these claims more thoroughly in a mixed population of patients, and to compare different computer programs for ECG analysis, LAS began last year to implement on the IBM-360 system several such programs now widely used in this country and abroad. This study is being conducted in cooperation with cardiologists at NHLI.

Progress during FY 71:

Programs from the Mayo Clinic and Public Health Service are now in regular operation on the 360 System, receiving ECG's from NHLI. The output from these programs is being systematically reviewed with Drs. Beiser and Itscoitz.

To facilitate study of cost-effectiveness, digitized EKG's are being stored on magnetic tape in categories: 1. Normal, 2. Borderline, 3. Hypertrophy, 4. Conduction Disturbance, 5. Infarct, 6. Supraventricular Arrhythmia, 7. Ventricular Arrhythmia, 8. ST-T Changes Alone, 9. Pediatric and 10. Miscellaneous (Axis Deviation Alone, etc.)

Analog-to-digital conversion is currently being done on the IBM-1827. However, to eliminate occasional errors in conversion which arise because of competition with other programs running on the 360, work is underway to switch this task to the LAS computing system, the MAC-16.

Proposed course:

The task of programming analog-to-digital conversion of ECG records on the MAC-16 will be completed during the first half of FY 72. A controlled, statistical evaluation and comparison of the Mayo Clinic and Public Health Service ECG analysis programs (and perhaps a third program if it can be implemented in time) will be undertaken with the Cardiology Branch. In addition to routine ECG diagnosis, these programs will be used by investigators at NHLI who intend to study a number of parameters, including ECG's, in patients with hyperlipidemia during a five year period of therapy (Dr. J. Brensicke, Lipid Branch, NHLI).

b) Cost-Effectiveness of Coronary Care Units

Background:

The Coronary Care Unit (CCU) represents a health care sub-system which may be amenable to a systems analysis approach. A need exists to determine optimal bed capacities for communities of different sizes and to develop a methodology for estimating the net costs and benefits accruing from CCU operations. Analytic techniques developed for the CCU will be directly applicable to other health care subsystems such as surgical recovery units, special diagnostic wards, respiratory care units, stroke wards, etc.

Progress during FY 71:

A deterministic cost-effectiveness model has been designed to relate the principal CCU parameters: mean arrival rate, mean service time, admissions policy, infarct fraction, partitioning of patient population and cost analysis. A computer program to simulate the stochastic aspects (related to queuing theory and involving inter-arrival and service-time distributions as well as specified bed capacity) has been constructed.

To compare results of the simulation program against actual operations, the charts of 525 patients who were admitted to the CCU of a local hospital have been reviewed. An additional data set on 695 patients admitted to a community hospital CCU in Dallas is nearly complete.

Proposed Course:

These patient data, codified by diagnosis on admission, age, length of stay, discharge status and other useful descriptors, will be used to validate or modify the computer simulation model. Once the model has been qualified, it may be used to estimate the results of varied admission and discharge policies, bed capacities, or other aspects of real CCU operation.

c) Computer Assisted Studies in Nuclear Medicine

Background:

The rapid expansion of the diagnostic use of radio-isotopes has been largely the result of two important advances: the invention of the gamma scintillation camera by Anger and the discovery of Technetium-99m, a radio-isotope with suitably low gamma energy for convenient counting, and short half life, reducing radiation exposure to the patient. Using Technetium-99m and the gamma camera, an excellent picture of the distribution of radio-isotope within an organ can be obtained in a matter of minutes. The application of established nuclear medicine techniques has greatly facilitated the early diagnosis of many diverse disease states, from brain tumors to joint inflammation. The next decade of advances in nuclear medicine will depend on computer technology. The gamma camera is capable of producing data at a high rate (20-30 kHz). Small dedicated computer systems have recently been developed which capture this data and hence allow important new opportunities for non-invasive study of pathologic anatomy and dynamic pathophysiology in a wide variety of disease states. In addition, these systems enable sophisticated off-line data processing by a large computer.

Progress during FY 71:

After one year of study and site visits, a joint LAS-CSL team has drawn up specifications for a small computer to be dedicated to the gamma cameras in the Nuclear Medicine Department of the Clinical Center. Contracting to buy the computer should be completed before the end of this fiscal year. Meanwhile, a program to estimate flow/volume ratio in an organ from the count density of the organ and of its arterial input is being developed. The program will have immediate application to compartments of the cerebral cortex and to renal flow. The data acquired by the gamma camera is presently stored on video tape. An abstract of work in progress is being prepared for the Journal of Nuclear Medicine.

Proposed Course:

When the small computer arrives and is checked out, it will be put to work immediately on the flow/volume problem to digitize and store counts and perform necessary arithmetic preliminary to further calculation as necessary, probably through the PDP-10. Other collaborative projects will begin as developed by I/D, Nuclear Medicine, and LAS staff.

Serial No. DCRT 3.3
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Methods in Cardiologic Research

Previous Serial Number: Same

Principal Investigators: J. J. Bailey
Assisted by: M. Horton, M. Douglas, W. Scott

Co-Investigators: M. Marcus, D. Glancy, S. Epstein, CB, NHLI
W. Whitehouse, W. Schuette, CC, BEI

Man Years: Professional - .75
Other - .5

Project Description:

A. Automated Angiographic Ventricular Volumes:

Background:

For the past several years, LAS, in cooperation with the Cardiology Branch and Surgery Branch, NHLI, has been developing and applying computer-based methods for determining the dynamic cardiovascular status of patients. These efforts have produced an extensive data acquisition and computer processing system for hemodynamic studies (DCRT Technical Report #2).

A new effort in this direction concerns the determination of ventricular volumes from angiographic data. At the Television Engineering Unit, CC, left ventricular cineangiograms (16mm) taken in the RAO position at 60 frames/sec are projected onto a plumicon television camera. An electronic video-tracking device simultaneously determines the area and the maximum length of the opacified chamber in each line frame; these data are recorded on magnetic tape. At the time the cineangiogram is made, the intraventricular pressure is recorded on magnetic tape along with a frame pulse generated from the fluoroscope.

Progress during FY 71:

The LAS contribution to this project began this year in the application of computer operations to the recovery of continuous digital volume and pressure information from the analog cineangiographic

data on magnetic tape. Pressure and volume tapes have been separately converted to digital form on the DCRT hybrid computer in Bldg. 10. Frame-by-frame matching of pressure and volume has been programmed for the 360 system. Noise generated by the plumicon television system is then minimized by a polynomial smoothing of the data. The 360 system produces CALCOMP plots of volume vs. time and pressure volume loops. When volumes determined by this automated method are compared with those obtained by manual planimetry, the average difference is 2-3%; Some bias exists in both methods however, since known volumes of simple test objects in the 20-360 cc volume range are accurate only to 6%.

Proposed Course:

From this system, which combines angiographic ventricular volumes with intraventricular pressure, continuous estimates of flow, stroke volume, stroke work, power curves, wall stress, ventricular compliance, and ventricular contractility will be obtained in patients with a variety of conditions including valvular defects and coronary disease. These data will be secured, routinely, while patients are undergoing cardiac catheterization. Knowledge and analysis of these parameters on a continuous basis through a cardiac cycle will promote better understanding of pathophysiology as well as more accurate quantitative diagnoses and prognoses for cardiac patients.

B. <u>Studies in Myocardial Contractility</u>:

Background:

Many cardiovascular studies evaluate the heart as a pump with three determinants: the preload, the afterload, and the performance of the heart as a muscle--referred to as "myocardial contractility", "inotropic state", "myocardial strength", etc.

Numerous attempts at finding a reliable index of myocardial contractility have revealed a number of promising candidates including: isometric contraction time, systolic ejection time fraction, mean systolic pressure, maximum change in pressure (max dp/dt) during isometric contraction, maximum dp/dt divided by the pressure (p) at which it occurs and maximal velocity of the contractile element (Vmax), (e.g., the Sonnenblick or Mason indices). From such parameters may be computed indices related to ventricular power or the efficiency of power conversion. Another important index of myocardial contractility may be derived from stress-strain relationships in the wall.

Progress during FY 71:

A review of the literature relating to myocardial contractility, heart wall stress and the indices noted above has been completed, and a mathematical evaluation and comparison of the various indices begun. In the meantime, since estimates of Vmax and (max dp/dt)/p are of considerable interest for use in drug studies and pre- and

post-operative evaluation of patients with idiopathic hypertrophic subaortic stenosis, a program has been written which can analyze the intraventricular pressure from thousands of beats collected on the patient in the cardiac catheterization laboratory. In patients with atrial fibrillation one can analyze the relationship of maximum dp/dt to end diastolic pressure, opening pressure, and RR interval for example, and test the validity of max dp/dt as an index of contractility.

Proposed Course:

Mathematical evaluation of proposed indices will continue and should lead to a method for computing the myocardial stress tensor at any point and at any time, given: wall thickness, radii, azimuthal angle, and intraventricular pressure. Then wall work and frictional losses can be computed as functions of wall stresses. When a model is programmed, it can be tested with data from the automated angiographic volumes (e.g., cardiac dimensions, volumes, output).

During the coming year, analog-to-digital conversion of pressure and volume tapes should be switched from the hybrid to the MAC-16 (the small LAS computer). Hopefully, this will permit data processing in sufficient time for presentation and study of CalComp plots at clinical conferences within the same day of collection. Such data will be available to LAS for further use in testing mathematical models of ventricular contractility.

Publications:

Marcus, M. L., Schuette, W., Whitehouse, W., Bailey, J. J., Glancy, D. L., and Epstein, S.). A completely automated video-tracking technique for the determination of dynamic changes in ventricular volume. Circulation (in press) (Abstract: XLII: Suppl. III, 101, Oct., 1970)

Serial No. DCRT 3.4
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Statistical Research in Clinical Pathology

Previous Serial Number: Same

Principal Investigator: Eugene K. Harris
Assisted by: David DeMets

Co-Investigators: E. Cotlove, Clinical Path. Dept., CC (deceased Sept., 1970)
D. Young, Clinical Path. Dept., CC
F. C. Bartter, W. J. Meyer, and others, EB, NHLI
G. Shakarji, DMB, DCRT

Man-Years: Professional - 1.5
Other - 1.0

Project Description:

Background:

Efforts during the past several years have been devoted to the study of analytic and biological components of variance in normal blood chemistries. The data base has consisted of 68 screened normal subjects, each providing 10-12 weekly blood samples, collected under controlled conditions and analyzed in duplicate for 15 common constituents. Source data for estimating long-term analytic variance have been provided by daily serum pool analyses during a two-year period.

Using standard statistical methods, analytic and biological variance components in blood chemistry tests have been estimated within and among normal individuals. To date, three reports on this work, with implications for laboratory analytic methods as well as interpretation of normal ranges have been published. These studies have shown (a) that where blood constituents under tight homeostatic regulation are concerned (e.g., sodium, magnesium, calcium), even the best current methods of analysis often mask small biological changes occurring in periodic controlled samples from a normal individual; (b) although very few of the tests studied (uric acid, cholesterol are exceptions) appear sufficiently powerful by themselves to distinguish among normals (i.e., to offer individual blood profiles), significant differences did appear among normal individuals with respect to mean levels in all constituents studied. Such differences were generally explainable only in part on grounds of age, sex or race, casting doubt on the general

usefulness of "normal ranges" specific for these variables.

Progress during FY 71:

Statistical research on methods of distinguishing biological and analytic variation within a single individual has been completed and published. As part of this research, a method was developed which permits estimating the variation from one individual to another with respect to intra-personal biologic variance, without having to separate analytic from biologic components of variance.

An experiment to uncover intra-individual biological variance by eliminating long-term analytic variance has been conducted and results are now awaiting publication. In this experiment, weekly samples were stored in frozen state and all analyses performed in one day at the end of the collection period. In a number of chemical constituents, particularly electrolytes, use of the most recent analytic techniques reduced the long-term analytic variance component to zero, and permitted recovery of small, but finite biological variances.

Further, this study presents evidence showing a remarkable diversity among normal individuals of the same age-class, sex, race and occupation with respect to mean level and intra-individual variability in the concentration of many common blood constituents. The implication is clear that such variations arise, in part at least, from unique, personal characteristics (e.g., diet, activity, genetic effects) rather than broad demographic factors.

A study of biological variations in serum ionized calcium, using the data provided by the 68 normal subjects, has been completed and submitted for publication. Based on application of the McLean-Hastings mass-action model for the direct estimation of Ca^{++} from total protein and total calcium, this study has yielded more comprehensive data on variations in Ca^{++} than heretofore existed. In many respects, results conform to recent estimates derived from calcium ion-selective electrodes. Intra-individual physiologic variation in Ca^{++} appears, on the average, to be extremely small, below the level of present analytic precision, even using electrodes.

Proposed Course:

Greater attention will be devoted to statistical studies of periodicity in clinical measurements, and the potential impact on drug effectiveness of circadian rhythms in susceptibility. In particular, statistical studies of physiological rhythms important in the evaluation and treatment of hypertension will be pursued. Associated studies will include the application of multivariate statistical methods to determine relationships between electrolytes in normal serum.

Publications:

1. Harris, E. K., Kanofsky, P., Shakarji, G., and Cotlove, E.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, II. Estimating biological components of variation. Clinical Chemistry. 16: 12 1022-1027, 1970.

2. Cotlove, E., Harris, E. K., and Williams, G. Z.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, III. Physiological and medical implications. Clinical Chemistry. 16: 12 1028-1032, 1970.

3. Harris, E. K.: Distinguishing physiologic variation from analytic variation. J. of Chronic Diseases. 23: 269-480, 1970.

4. Buchsbaum, M., and Harris, E. K.: Diurnal variation in serum and urine electrolytes. J. of Applied Physiology. 30: 1 27-35, January, 1971.

5. Young, D. S., Harris, E. K., and Cotlove, E.: Biological and analytic components of variation in long-term studies of serum constituents in normal subjects, IV. Results of a study designed to eliminate long-term analytic deviations, Clinical Chemistry. 17:5, May, 1971.

Serial No. DCRT 3.5
1. Laboratory of Applied Studies
2. Applied Mathematics
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Applied Mathematics Unit

Previous Serial Number: Same

Principal Investigator: J. Fletcher
Assisted by: J. Ashbrook, E. Hill, K. Drew

Co-Investigators: W. Rall, AMR, NIAMD; R. Mejia, LAS, DCRT; A. Spector, Univ. of Iowa (formerly with Lab. of Metabolism, NHLI, under D. Steinberg); J. Gonzalez-Fernandez, AMR, NIAMD; M. Magar, Naval Medical Research Laboratory; E. Evarts, LNP, NIMH; B. Hubbard, Univ. of Maryland

Man Years: Professional - 4.0
Other - 0.5

Project Descriptions:

Background and Objectives:

Established within the Laboratory of Applied Studies during the preceding reporting year, the primary responsibility of the Applied Mathematics Unit is to provide NIH scientists with mathematical competence for bio-mathematical modeling and data analysis. This competence includes both theoretical and applied techniques, as well as numerical computation methods. Each individual in the unit has a primary specialty in computer science or mathematics, and each is a capable computer programmer.

Progress during FY 71 and proposed course:

1. Macromolecule-Ligand Binding:

One facet of this project, studies of the mathematical equivalence of binding models (comparing the stepwise series of equilibrium reactions with the Scatchard model) has been completed and published. However, it has been shown that the parameters of the Scatchard model are not biologically meaningful unless certain restrictive assumptions are met. On the other hand, analysis of macromolecule-ligand interactions by stepwise equilibria is sufficient generally to account for cooperativity and other phenomena that may be associated with free fatty acid-protein interaction.

As a result of this research, it is desirable to reanalyze earlier data. Initial results of such reanalysis are being presented at the 62nd Annual Meeting of the American Society of Biological Chemists. A new study has been initiated to investigate the binding of medium-chain fatty acids, important in chemotherapy studies, to human serum albumin. The effects of error in various types of data and with various models are being investigated so that other researchers may be advised of appropriate models for analysis of their data. Parallel studies concern the resolvability of a model's parameters from a given set of experimental data.

2. Applications of Pure Boundary Value Problems in Parabolic Partial Differential Equations to the Modeling of Biological Diffusion Processes:

   This is a newly begun study to review and develop mathematical techniques needed to solve linear and nonlinear diffusion models. After a review of the literature, a model for unsteady flow and gaseous exchange in the micro-circulation has been derived. The mathematical properties of this model are under study and numerical solutions are being constructed. Two numerical methods, a fractional step method developed by B. Hubbard, Univ. of Maryland and the older Crank-Nicholson procedure, are currently being debugged. The objective is the ultimate capability of simulating the transport-diffusion system of the capillary-tissue arrangement in muscle.

3. On-Line Modeling System:

   This project, completed last year, has been described and published during FY 71 as a DCRT Technical Report (#6). Briefly, an interactive computer display system, designed and implemented on the PDP-10 computer, allows users to specify mathematical models, manipulate data files, change parameter values and view results in real time.

4. Simulation of the Utilization of a Coronary Care Unit:

   In cooperation with the Medical Systems Unit (Project 3.2), a simulation model of a coronary care unit has been designed which uses linked lists to study the utilization of a coronary care unit with a fixed number of beds as a controlled variable. The model has been programmed in a new simulation language called GASP II, a language for describing the behavior of dynamic systems. Variables in this model are the admission and discharge (length-of-stay) policies and the number of available beds. Statistics from existing care units are being collected for comparison with results of the model.

5. General Support Activities:

   1. Calma to Curve Fitter--A flexible tape reading technique was developed so that programs for generalized least squares model

fitting (DCRT Reports Nos. 1, 5), can utilize data obtained directly from the CALMA digitizer system. The CALMA digitizer can convert analog data on paper or microfilm to digital information on magnetic tape. These programs are now standard and are being employed by a number of NIH scientists.

2. CalComp--Extensive revisions were made to the CalComp Manual to clarify, correct, and expand the documentation. These have been issued as revisions to DCRT Technical Report #3, CalComp Manual. Consultation and advice is now provided to the users of the CalComp plotter as needed. A member of the Applied Mathematics Unit is currently in charge of maintaining and updating CalComp software and documentation, and conducting a DCRT Training Course entitled, "Introduction to the CalComp Plotter."

3. Part of a set of Numerical Analysis Utility Programs written at the Oak Ridge National Laboratory have been tested and are now running on the PDP-10 computer in DCRT. (e.g., Matrix inversion, solution of linear systems equations, integration procedures, etc.)

Publications:

1. Fletcher, J. E., Spector, A. A., and Ashbrook, J. D.: Analysis of Macro-molecule-Ligand Binding by Determination of Stepwise Equilibrium Constants. Biochemistry. 9:4580-4587, 1970.

2. Magar, M. E., Steiner, R. F., and Fletcher, J. E.: Analysis of Protein Ligand Equilibria. Journal of Theoretical Biology. (in press).

3. Spector, A. A., Fletcher, J. E., and Ashbrook, J. D.: Analysis of Long-Chain Free Fatty Acid Binding to Bovine Serum Albumin by Determination of Stepwise Equilibrium Constants. Presented at the 1971 American Society of Biological Chemists meetings in San Francisco, June 18, 1971.

4. Hill, E.: The On Line Modeling System, Parts I, II, III. DCRT Technical Report #6, April, 1971.

5. Minicozzi, W., and Stroot, M.: On the Determination of Interaction Energy Functions, II. Crystalline Formic Acid. Journal of Computational Physics. August, 1970.

6. Rinzel, J.: A Mathematical and Numerical Treatment of Neuronal Dendritic Spines. Presented at SIAM National Meetings, July, 1970 in Denver, Colorado. (a paper with W. Rall, NIAMD in preparation).

Serial No. DCRT 3.6
1. Laboratory of Applied Studies
2. Office of the Chief
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Simulation and Modelling in Neurophysiological Research

Previous Serial Number: None

Principal Investigators: Erik Pottala (Project A)
James Mortimer (Project B)

Co-Investigators: T. R. Colburn, IR, NIMH (Project A)
E. V. Evarts, LNP, NIMH (Project B)

Man-Years: Professional - 2.2
Other - 0.1

A. Hardware Neural Modelling

Background:

The objective of the hardware neural modelling study is to investigate small neural nets through the use of physiologically realistic hardware neural models which incorporate a distributed input system (analogous to a dendritic net) and are able to simulate action potentials.

Most previous investigators of neural nets have relied on point neural models. Although these point models typically used rectangular voltage pulses for their inputs and action potential, they were capable of exhibiting several features of physiological neurons, e.g., temporal summation of inputs, absolute and relative refractoriness, and adaptation. They could not simulate the spatial summation of inputs or the electrotonic spread of an action potential or postsynaptic potentials.

Progress during FY 71:

A new hardware model has been developed which overcomes these shortcomings and displays most of the characteristics of a physiological neuron. Among these properties are the model's "synaptic" inputs which are actual conductance changes and the model's simulated action potential, created through depolarizing and hyperpolarizing conductance changes.

The model is currently being used to study input-output, steady state firing characteristics for a single neuron; i.e., to show the relationship between the model's firing frequency and stimulus intensity, position, and frequency.

Proposed Course:

The model neuron will first be used to investigate the effects of the direction of applied stimulus (i.e., distal followed by proximal stimulation, or vice versa) on the post-stimulus-time histogram of firings. To facilitate these experiments, an interactive data acquisition and control system is being implemented on the LAS computer (the MAC-16) and will be perfected during the first half of the coming year.

B. A Functional Model and Experimental Study of Mammalian Cerebellar Cortex

Background:

The objectives of this project are:

1) To evaluate several alternative hypotheses of cerebellar function through study of the electrophysiological activity of the cerebellar cortex of a monkey during the performance of motor tasks;

2) To relate the physiological findings of this study to known cerebellar circuitry by means of a computer simulation method, already implemented;

3) To develop a new functional model of mammalian cerebellar cortex suitable for computer simulation.

The study of large and complex neural networks requires interdisciplinary communication between physiologist and modeler. In the past, this communication has been inadequate; with few exceptions, the modelers have developed unrealistic and oversimplified models, while the experiments designed by the physiologists have not dealt with the theoretical issues raised by the modelers. In the present project this lack of communication is attacked through an interdisciplinary study under the direction of the principal investigator. Two main areas of research are involved:

1) Design and execution of neurophysiological experiments, which will provide the kind of data required for the construction of a functional model of the cortex, and

2) The development of useful and realistic models of the cerebellum. Computer methods are employed extensively in both of these areas.

Previous work of the principal investigator has focused upon the development of a computer-simulated model of mammalian cerebellar cortex. Representation of the cortical network as a cellular automaton has made it possible to deal with much larger and more realistically-structured networks, than had been examined with earlier computer models.[1] Several experimentally-testable predictions have been generated by this model. For example, it is predicted that the presence of a particular population of cortical elements, the Purkinje recurrent collaterals, should lead to the synchronization of the firing patterns of Purkinje cells situated in

adjacent transverse rows. This prediction can be verified by comparing the Purkinje cell activity in animals which have recurrent collaterals with that in animals lacking these neuronal elements. Another prediction amenable to immediate experimental verification, is the remarkable difference in the time course of Purkinje activity predicted under various conditions for the preparation (anesthetized, decerebrate, unanesthetized) for certain classes of stimuli. If the predictions of the simulation model are correct, the function of the cortex might differ considerably under various states of cortical arousal.

Progress during FY 71:

This year has been devoted to the design and initial testing of an experimental protocol for obtaining data from cerebellar neurons during the performance of relatively complex motor tasks. One such experiment, being conducted in cooperation with the Laboratory of Neurophysiology, NIMH, involves predictive and nonpredictive eye-hand tracking and its electrophysiological correlates in the monkey.

In conjunction with this experiment, the principal investigator has designed and constructed a special-purpose digital computer, which has complete control over the presentation of the stimulus, evaluation of the response, and data collection. The investigator is therefore free to concentrate on those aspects of the experiment not easily amenable to automation, e.g., the positioning of the microelectrodes.

Proposed Course:

The experimental procedures outlined above will be perfected, data obtained and analyzed particularly with respect to the relationships between simple and complex Purkinje spike activity and motor behavior. Modifications to the computer simulation model and further computer experiments will follow.

1/ J. A. Mortimer: "Large Network Models: State-of-the-Art and Expectations", Winter Conference on Brain Research, Aspen, Colorado, January 17, 1971

July 1, 1970 through June 30, 1971

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT - 4
Serial Number

2. COMPUTER SYSTEMS LABORATORY

3. Alan Demmerle
Chief

The Computer Systems Laboratory (CSL) concentrates on data system problems that do not lend themselves to solution through use of DCRT's central computer facility and require instead, the development of peripheral computer systems or special purpose hardware. CSL designs and implements computer based data systems for any promising application at NIH, even though during our first few years we have concentrated on the areas of laboratory automation and the delivery of health care.

A specific area which promises potential benefit by the application of the computer technology is identified; the way in which the computer is to be used is determined, the hardware and software aspects of the system are designed and implemented, and the users instructed in the use of the system. Such work requires a multidisciplinary group effort and requires two to four years from inception to completion.

Because of the large number and diversity of experiments and instruments and the unique combinations of requirements, NIH represents a unique environment for the application of computers. There has been no experience in the application of computers in such an environment, so at the outset, one can only speculate as to the potential benefits of "computerization." We have therefore engaged in what amounts to experiments to determine the benefits of computers in several different kinds of research laboratory and medical care environments. These trials, to determine whether or not the computer will really facilitate better research, involve a sizeable investment in capital and development manpower.

Within the next year or two some of the systems which we have developed will be operating sufficiently smoothly to assess their value. Only then can the true cost-benefit-relationship for these systems be determined. A lengthy period is necessary, not only because of the size of the development effort but also because it takes time for the scientists to get accustomed to using the system. In some cases they may need to adjust their experimental procedures and the way they collect and use data to get the maximum benefit from their new tool, the computer. NIH is providing leadership and innovation in this type of computer application; in evidence of which, the U.S. Geologic Survey, Merck, Sharpe & Dohme and the FDA, among others, have been here to look over some of these computer systems and are considering copying them.

Computers in the Laboratory

The laboratory automation work, to date, falls in two categories: the "computerization" of biochemistry laboratories, and the "computerization" of psychological studies. In the first category there is the computer system for NIDR, one for NIAMD, one for the NHLI's Mass Spectrometers, one for the NICHD's Gerontology Research Center in Baltimore, and one for use with the PSL-DCRT NMR Spectrometer.

Although the systems are in many ways different, they are all used to collect data directly from instruments, to calibrate, to format, to perform calculation on the data, and to produce analyzed experimental results for the bench scientist, during, or immediately after, his experiment. The processed data is then available in a form that should be easy to use by the experimenter, for example, as plotted results or graphical results displayed on a CRT. Generally, the system collects data from several instruments simultaneously so the marginal cost per user, or per experiment, is minimized. The systems provide the scientist with analyzed data at the time he is conducting his experiment, thus allowing him to modify his experiment while it is still set up and his experimental samples are still available. It allows him to do some experiments that he could not do before, for example, stimulus-response experiments using implanted micro electrodes can be performed, the stimulus changed and more data collected and analyzed before the cell in which the electrode is implanted dies. These computers also allow him to analyze more data in more complex ways than was ever possible before. The state of progress of these systems is reported in the individual progress reports 4-1, 4-2, 4-3, 4-4, 4-5.

In the second category is the NIMH Building 10 system to be used to facilitate psychological studies. In these studies the patients interact with the computer. Here also, the computer is shared among a number of concurrent studies --each scientist can use the computer simultaneously, yet independently, of the others. The system serves to collect data, control and modify game situations and control and format displays of various kinds.

Computers in the Delivery of Health Care

There are four projects in this Laboratory relating to the delivery of health care. The first one (4-7) is an effort to bring easy to use, economical and reliable computer services to the practicing physician. He will be able to use a ubiquitous terminal such as a standard telephone, to access commercial computer systems for assistance in such activities as diagnosis and therapy planning. This system is designed with the community physician in mind, as well as NIH clinical requirements. The second and third projects relate to the care of critically ill patients. In these two projects (4-8 & 4-11) we have concentrated on using the computer in cardiovascular studies with NHLI, and in working toward a computer system to assist in an intensive care environment. The fourth project (4-9) is related to the automatic processing of medical English text, and is viewed as the only feasible way to make the vast quantities of textual research data, such as autopsy and surgery reports, available for analysis and the testing of hypothesis. The details of progress in these projects can be found in the Individual Project reports.

Most of these projects began prior to this reporting year. Most represent collaborative work with other Institutes. Attached is a table which shows, graphically, the CSL personnel effort put into these projects over the last 3 years, the capital expenditure required for these projects and the source of these funds. The capital expenditures portion of the table has one entry for equipment which includes both commercially purchased equipment and the cost of components for the equipment developed in this Laboratory, and one entry for central facility computer charges expended on the project for computer program development. The personnel effort includes that required for problem analysis, system engineering, equipment design engineering and programming, all of which is done by professionals in the field of engineering, mathematics, physics, chemistry, linguistics, medicine and computer science.

		CSL MANPOWER (man-years)			CAPITAL Equipment			
Project #	Name	FY-69	FY-70	FY-71	Source	Amount	FY	Computer Charges
DCRT 4-1	NIDR	5	5	2	NIDR	100 (K$)	68	
					DCRT	150	69	
					DCRT	0	70	12(K$)
						0	71	1
4-2	NIAMD	5	5	4	NIAMD	180*	68	
					DCRT	25	69	
					DCRT	20	70	3
					DCRT	2	71	6
4-3	Mass Spec.	2	2	2	DCRT	76	69	
					DCRT	10	70	1/2
					DCRT	2	71	1
4-4	Gerontology	1	3	2	NICHD	111	69	
						33**	70	0
						0	71	0
4-5	NMR	0	1/2	1 1/2	DCRT	68**	70	0
4-6	NIMH	1/2	1	4	NIMH	243*;**	70	0
							71	5
4-7	Med. Telecomm.	6	6	6	DCRT	125	69	
					DCRT	112	70	12
						16	71	29
4-8	Cardio-Vascular	3	5	4 1/2	DCRT	20	69	
					DCRT	21	70	1
					DCRT	11	71	2
4-9	Auto. Processing of Med. English	4	4	4	DCRT	0	70	30
							71	23
4-10	Consultation	2	1 1/2	1	DCRT	0	71	0
4-11	Intensive Care Unit	0	0	1	DCRT	325	71	0

*This is the purchase price of the system which was partially purchased and partially leased.

**The figures listed in last years Annual Report have been corrected in this years report. Discrepancy due to actual expenditures after FY-70 report had been prepared--previous figures had been estimates.

(Cost figures are approximate.)

Serial No. DCRT 4-1
Computer Systems Laboratory
Systems Design
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer System for NIDR

Project Leader: Daniel Syed

Objectives:

The use of the computer as an adjunct to laboratory procedures in the National Institute of Dental Research is intended to both augment and expedite the biochemistry and neurophysiology research programs of that Institute. Specifically, the computer system will modernize techniques for acquiring data from laboratory instruments and will provide for the real-time analysis of data from selected experiments. Ultimately, it is planned to use the computer to control experiments in real-time.

Background:

Subsequent to a detailed requirement study started in February 1967, a process control class computer system was designed and specifications generated. Open competition procurement resulted in the selection of a Honeywell DDP-516 computer which was purchased in June 1968 and installed in the Dental Institute, Building 30, in July 1969.

FY-71 Activities:

During the past year Honeywell supplied system software was modified to permit operation in a multiprogrammed mode. Specifically, the system is now capable of acquiring and analyzing data acquired simultaneously from amino acid analyzers, scintillation counters and neurophysiological experiments. In addition to the generation of applications programs, user oriented executives have been implemented for use with neurophysiological projects and experiments utilizing scintillation counters. An experimental link to the central PDP-10 computer has been completed.

Future Efforts:

Efforts during the coming year will be concentrated on interfacing a spectrophotometer, an x-ray diffractometer, gas chromatographs and a ph meter to the Honeywell 516 computer. Implementation of the communication link to the central facility IBM 360 computers, delayed by hardware reliability and core storage problems, is also planned. Continuing efforts will be required to simplify use of the system by the research investigators.

Serial No. DCRT 4-2
Computer Systems Laboratory
Laboratory Applications
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer System for NIAMD

Project Leader: Marvin Shapiro

Objectives:

The NIAMD computer located in Room 110, Building 2, is a real-time data acquisition system which has been in operation since March 1970. It is being used as a research tool by scientists in a number of ways. Data collection is done quickly and accurately with the computer and is then available in a form suitable for further processing. This represents a significant reduction in the time required to complete calculations and to plot results over previous methods. These improvements, plus the ability to use the computer for feedback control of some experiments, should result in a significant change and improvement in the experimental procedures being used by many scientists in NIAMD.

Background:

A computer system for real-time data acquisition was designed in early 1968 and a system meeting the requirements (a Honeywell DDP-516 computer) was ordered in June of 1968. The system was delivered in July 1969. NIH work on interfacing instruments to the computer and on software modifications to the monitor supplied by Honeywell began in July 1968. The system became operational for routine data collection in March 1970, at which time four instruments were interfaced to the computer.

FY-71 Activities

Two more instruments, a Computer of Average Transients (CAT) and an analytical ultracentrifuge, were interfaced to the computer and incorporated into the software operating system. Two packages of software for the plotter were completed. The computer began to be used steadily for data acquisition, particularly with the Cary 14 and IR-7 spectrophotometers. The number of data acquisition users increased from 5 to 10.

Future Efforts:

During the coming year work to be done will include the following:
(1) Interfacing of more instruments, including a CARY 15 spectrophotometer.
(2) Completion of a connection to two small computers, a Raytheon 704 located in Dr. Hagins' laboratory and a PDP-8 located in Dr. Davies' laboratory will thus receive support in the form of added computational power and backup storage.
(3) Connection to the central NIH computer system via telephone lines. This project has been delayed by the necessity for optimization of existing programs in order to release additional core storage.
(4) The development of more application software for individual instruments, notably the ultracentrifuge.

Serial No. DCRT 4-3
Computer Systems Laboratory
Laboratory Applications
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: NHLI Mass Spectrometer Computer System

Project Leader: Marvin Shapiro

Objectives:

The National Heart and Lung Institute Mass Spectrometer Computer System located in Dr. Fales' laboratory is designed to collect and process high and low resolution mass spectrometer data. The output from the system is used in the elucidation of the structure of complex molecules. The computer provides a much higher level of performance than was previously obtainable with the mass spectrometer system alone. Complete mass assignments for over 200 masses can now be done, to 6 digit accuracy, in minutes, as compared to many hours of tedious hand work without the aid of the computer. Also, the digitization of the data provided by the computer system enables further processing to be done on the data on a larger computer.

Background:

The MS-9 high resolution and the LKB low resolution mass spectrometers were already available for interfacing to a computer. The initial version of the computer system, consisting of a PDP-8I computer with 4K of memory, a teletype, and A/D conversion equipment, was delivered and interfaced to the MS-9 in December 1968. The incremental plotter was delivered in February 1969 and a high-speed paper tape reader-punch, an incremental tape and a disk were added in June 1969, to complete delivery of the system. The LKB low resolution spectrometer was connected to the computer in February 1970. In March 1970 the system software called SERF, which utilizes the disk, was delivered.

FY-71 Activities:

Chiefly, four projects were worked on:

(1) The system was used on a routine basis to obtain medium resolution mass spectra of compounds, utilizing a computer data acquisition system provided by Dr. Markey of the University of Colorado Medical Center.

(2) The connection of the computer to the LKB low resolution mass spectrometer was completed and the supporting software is being developed.

(3) A program which uses low resolution mass spectrometer data to identify poisons and barbiturates was completed and is used for emergency overdose cases.

(4) A large amount of software for the plotter and for system improvements was completed.

Future Efforts:

The system has not yet been accepted and paid for, primarily because of deficiencies in the DS20 software system. We are awaiting the DS30 system which supposedly corrects all the major problems.

The system will be improved to provide for routine collection of high and low resolution mass spectrometer data. It is hoped that the system will be fully operational in the coming year and that a typical week might see a throughput of as many as 20 high resolution samples (as opposed to the past rate of less than one sample) per week.

Work will continue on using the mass spectrometer for identifying amino acid PTH derivatives which result from an Edman degradation of a protein. A new computer instrument system, which includes a small computer and a quadrupole mass spectrometer, is being purchased for this application.

Serial No. DCRT 4-4
Computer Systems Laboratory
Project Development
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer System for Gerontology Research Center

Project Leader: Perry S. Plexico

Objectives:

The computer system at the Gerontology Research Center (GRC), NICHD, in Baltimore, is used to accommodate various on-line control and data acquisition experiments and off-line data processing tasks in support of research by the Center into the nature of the ageing process. The research efforts are multidisciplinary and encompass the techniques of biochemistry, molecular biology, physiology, and psychology; hence, the applications for which the computer is used are widely varied.

Background:

In February 1968, the GRC requested that the Computer Systems Laboratory, DCRT provide assistance in the selection of a computer system meeting GRC research requirements. A study of GRC research needs was undertaken and resulted in the preparation of detailed computer system specifications which were used in March 1969 to solicit proposals from computer manufacturers. The computer system was procured from Raytheon in June 1969. It was delivered in February 1970 and was accepted by the Government in May 1970.

FY-71 Activities:

The system has been used throughout the past year by members of the GRC scientific staff for off-line processing and for on-line psychological testing and problem solving experiments. During the same period, members of the CSL staff have continued development of software to automate a glucose/insulin infusion control experiment. This is a closed loop control experiment in which the computer controls infusion of glucose into the bloodstream of a subject while monitoring blood glucose level, the objective being to achieve a better understanding of the underlying causes of diabetes and the effects of ageing on glucose tolerance. Although software to support this project was to have been completed during the fall of 1970, its development time was extended by unexpected difficulties in automating the experiment, by instrumentation changes, and by changes to the GRC experimental protocol. All programming has now been completed with only final testing with an actual experiment left to be performed.

Future Efforts:

Upon completion of software for the glucose/infusion experiment, it is planned that involvement of CSL personnel with this system be terminated.

Serial No. DCRT 4-5
Computer Systems Laboratory
Processor Design
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Aided Nuclear Magnetic Resonance Spectroscopy

Project Leader: Arthur Schultz

Objectives:

The Physical Sciences Laboratory, DCRT, in collaboration with the NIAMD, plans to use a pulsed nuclear magnetic resonance (NMR) spectrometer in conjunction with a small scale on-line computer system for the performance of high resolution Fourier transform spectroscopy. The advantage of Fourier transform spectroscopy, as compared to conventional frequency sweep techniques, is that significant improvement in signal-to-noise can be realized in appreciably less time.

Background:

As a result of the Computer Systems Laboratory (CSL), DCRT system study on NMR Fourier transform spectroscopy, a system was configured, proposed, and procured in June 1970. During the initial phase of NMR experimentation, proton and carbon 13 resonance will be investigated.

FY-71 Activities:

A fixed price contract was awarded to Raytheon Company in June 1970 for a R-704 with 12K core, fixed head disk, 9-track magnetic tape and laboratory interface. Delivery of all items is now complete and the system is undergoing performance testing. Additional hardware has been fabricated by CSL for the instrument interface. These include: (1) a programmable clock used for selecting appropriate data sample rates, (2) synchronization control between the computer and pulsed equipment, and (3) necessary data communication hardware.

This project involves not only two Laboratories in DCRT, but also BEIB, in DRS. The project is currently awaiting delivery of the pulsed radio frequency instrument from BEIB, DRS.

Future Efforts:

With regard to the computer aspects of this project, work will continue on the development of the software system to accommodate data acquisition, signal averaging, Fourier transform and data display. These computer related aspects should be complete in about 6 months. We also expect to add a hard copy facility to make the system more versatile and somewhat easier to use.

Serial No. DCRT 4-6
Computer Systems Laboratory
Systems Design
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer System for NIMH

Project Leader: Victor Colburn

Objectives:

This project involves implementation of a real-time computer system for support of research studies which are being conducted in the National Institute of Mental Health at NIH. The system will provide on-line data acquisition, analysis and control for experiments in Problem-Solving, Learning, Evoked Response, Perception, Family Interaction, and others. It will also support investigations which relate electroencephalographs and other physiological variables to clinical diagnosis and classification of NIMH patients. It will also provide certain off-line support for these same scientific activities.

Background:

CSL conducted studies of these NIMH requirements during calendar 1967. Funding problems limited initial procurement to 70% of intended total system. A contract was signed in November 1969 with Systems Engineering Laboratories of Florida. NIMH has some computer support personnel, both programmers and engineers, so the work required to implement this system was divided between CSL and NIMH. CSL's responsibility was to implement and extend the manufacturers Real-Time Executive (RTX) system to best fit the NIMH multi-user environment.

FY-71 Activities:

The computer was delivered to NIH, Building 10, on June 30, 1970. The RTX operating system was installed and checked out by CSL in August. The overall system was officially accepted by NIH in October. Remote teletypes were installed in several experimenter's areas and users commenced with time-share operation.

A USERS GUIDE was written by CSL and distributed to NIMH in December 1970. CSL wrote and documented system programs for: (1) a multi-user interval timer control; (2) a multi-user disk data file access; (3) a system interface program for experimenters test panel; and (4) several other general purpose utility routines.

Test runs of real-time experiments by two NIMH scientists began in early March 1971. At the end of FY-71, five NIMH scientific users are involved, to varying degrees, with real-time operations.

Future Efforts:

The system, as delivered, was extensive and flexible, but was not inherently a multi-user system. CSL will continue to tailor the software system to the NIMH multi-user environment. This effort will be guided by two basic objectives: (1) to make the system more convenient for the scientist to use, and (2) to provide maximum protection from user interaction.

Serial No. DCRT 4-7
Computer Systems Laboratory
Project Development
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Medical Telecommunications

Project Leader: Perry S. Plexico

Objectives:

The objectives of this project are: (1) to facilitate use of the computer for the practicing clinician, medical research physician and supporting paramedical personnel by developing convenient economical and reliable methods for remote access to computer-stored information and computations, and (2) to develop useful computer programs to assist the health professional in patient-care decision-making.

Background:

With the advent of time shared computer systems during the last five years, the potential for bringing computer technology to bear on some of the problems of health care delivery has become economically feasible. Unfortunately, however, the terminal devices which must be used with commercial time sharing services are not well suited to clinical applications due to their high cost, large size, and lack of wide availability. Furthermore, each commercial time sharing system has its own idiosyncracies of input and output, thus making its use by the non-computer oriented physician unattractive. Finally, no single commercial system is adequately reliable. These factors led to the exploration of the use of the telephone as a computer terminal, which, in turn, resulted in the procurement, in the Spring of 1969, of a small computer system for use as a communications processor. This system has been designed to translate between each commercial time sharing system's format to one familiar to health personnel, and to allow for automatic switching from one commercial system to another so as to provide adequate reliability even when any given time sharing system is inoperative. Inquiries are accepted from a standard pushbutton telephone, and voice responses are provided to the user over the same telephone via audio response units developed by CSL.

FY-71 Activities:

During FY 1971 development of the first phase system software to accommodate touchtone telephone input, voice response, and communication with commercial time sharing services for four simultaneous users was essentially completed. The system can be expanded to accommodate more simultaneous users, when necessary. Additional auxiliary communications equipment to permit automatic dialing of commercial time sharing services was also developed. The system's ability to communicate with the DCRT PDP-10 time sharing system and to satisfactorily execute programs from a terminal consisting only of a remote touchtone telephone was demonstrated. Software was developed for the management, generation, and modification of the system vocabulary for voice answerback.

Future Efforts:

During the summer of 1971 a field trial of the Medical Telecommunications system is planned during which knowledgeable members of the medical community will use the system and comment upon its applicability, as well as suggest additional useful applications programs. The exact direction of future developments, e.g. whether effort should be concentrated on additional applications programs or more versatile communications software, etc., will be conditioned by this trial.

Serial No. DCRT 4-8
Computer Systems Laboratory
Systems Design
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Cardiovascular Studies

Project Leader: Daniel Syed

Objectives:

The beat-by-beat analysis of electrocardiographic and blood pressure waveforms can provide great insight into the pathophysiology of the cardiovascular system. In collaboration with the Cardiology and Surgical Branches of NHLI, this project attempts to monitor and analyze these physiological variables through the use of recent developments in digital computer technology.

Background:

Previous CSL efforts center around the development of arrhythmia monitoring systems in the form of real-time software and special purpose hardware preprocessors. A hardware aortic blood pressure preprocessor has also been completed. The hardware preprocessors utilize a novel optical isolation circuit which insures the safety of the patient being monitored. A network of FM datasets has been implemented to allow transmission of biomedical waveforms from the heart surgical suite, post operative recovery room, and other patient care areas to the hybrid computer facility in Building 10 or the CSL electronics laboratory in Building 12A. Software systems have been implemented for the quantification of myocardial function by analysis of the motion of the heart border and the Mason Index analysis of the left ventricular pressure waveform.

FY-71 Activities:

During FY 1971 the arrhythmia monitoring software was heavily utilized in the evaluation of anti-arrhythmic chemotherapy involving the coupled use of Xylocaine and Atropine. Additional software was completed to allow the measurement of S-T segment displacements during myocardial ischemia. The Mason Index software package was expanded and utilized to study changes in myocardial contractility due to the introduction of a coronary artery bypass graft. Several techniques for the determination of Mason Index parameters were compared.

An experimental study was undertaken in the animal laboratory which provided data to allow the comparison of the several arterial pulse contour methods for the estimation of stroke volume. An ultrasonic doppler flowmeter is being utilized in an investigation to gain insight into the changes in peripheral resistance which degrade the accuracy of the pulse contour methods.

A hardware ECG preprocessor designed by CSL has been refined and clinical evaluation of the device begun. A simple digital algorithm for the detection of premature ventricular contractions is under evaluation. Development of a battery operated FM acoustic coupler, which acquired an ECG from a patient's

fingertips is being developed. Optical isolation techniques have been implemented in an in-line coupler which separates a subject from any standard ECG amplifier, providing complete electrical isolation.

Future Efforts:

Continued analysis of clinical and animal laboratory data is planned. Many of the developments of this project will be modified for use with the Clinical Center intensive care computer system (Project 4-11) when that system becomes operational.

Serial No. DCRT 4-9
Computer Systems Laboratory
Medical Information Science
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Automated Processing of Medical English

Project Leader: Milos Pacak

Objectives:

The major objective of the project is the completion of a fully automated system for machine encoding (automatic indexing) of medical diagnoses.

Background:

The project for automated encoding of medical diagnoses was started three years ago. It includes the acquisition of textual information, interrogation of the SNOP dictionary, and the necessary grammar and logic to allow for identification of the information content of medical diagnoses, and storing of relevant information in the data file.

FY-71 Activities:

During the past year the main effort was concentrated on further improvements of the system for automated indexing and minor revisions of the SNOP in a close cooperation with the Department of Pathology. We have completed the encoding of about 1,800 medical diagnoses which will be entered in the data file. The present system encodes correctly approximately 85% of the surgical pathology data presented in the form of unrestricted medical English.

The work is being continued on further development of the procedure for the morphosemantic segmentation of compound medical terms and their semantic interpretation. At present, we have analyzed approximately 1,200 semantic constituents.

The design of a formal semantic model for medical English is under further development.

Future Efforts:

1. Encoding of medical diagnoses for 1971 and for the years prior to 1968.
2. Testing of a program for morphosemantic segmentation and interpretation.
3. Preparation of an experimental program for testing the validity of several hypotheses for the development of a formal semantic model for medical English.

Serial No. DCRT 4-10
Computer Systems Laboratory
NA
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Systems Laboratory Consultation

Project Leader: Alan Demmerle

Objectives:

This project incorporates a number of consultation activities directed toward providing requesting Institutes with guidance in any aspect of computer systems design and use.

Background:

The project seeks to afford assistance to the intramural and extramural programs of all Institutes, however, to date, consultative activities have been directed primarily toward the support of the data management portions of large dollar value, long term contracts and grants sponsored by the NHLI. During collaboration with NHLI, efforts have included providing contractor guidance, designing a complete system or simply offering advice to NHLI contract office.

FY-71 Activities:

During the past year effort was concentrated on the support of two major NHLI projects: the Myocardial Infarction Program and the Blood Resources Program.

Future Activities:

Future efforts in this project are determined by ad hoc requests from the Institutes.

Serial No. DCRT 4-11
Computer Systems Laboratory
Systems Design
PHS-NIH Bethesda

Project Report
July 1, 1970 through June 30, 1971

Project Title: Intensive Care Unit Computer System

Project Leader: Daniel Syed

Objectives:

This project is intended to develop a computer system, for intensive care monitoring, that is capable of performing the following functions: (1) vital sign monitoring with feedback to clinicians and nursing staff, (2) statistical analyses, trend identification and other analyses performed on historical data, and (3) eventual computer control of selected patient variables. Performed in collaboration with the Surgical and Cardiology branches of NHLI, on-line data will be obtained from the post operative recovery area of NHLI, from the Clinical Center Intensive Care Unit and from NHLI Catheter Laboratories.

Background:

In 1969 members of the CSL staff designed a computer system to be used in intensive care patient monitoring at five Universities funded for Myocardial Infarction Research by NHLI. It is planned to modify this system for the NIH Clinical Center environment.

FY-71 Activities:

Procurement of the computer system and preparation of the site was initiated. The development of first phase system goals was also begun.

Future Efforts:

The computer system, which is expected to arrive in DCRT by August 1971, but not installed in Building 10 until the Spring of 1972, must be made operational with an initial complement of physiological variables such as EKG, blood pressures, etc. This statement implies a comprehensive range of activities including applications and system software generation, signal transmission, signal interfacing, information storage and retrieval, etc. The system must be developed to accommodate respiratory and blood gas parameters and to ultimately support control of appropriate variables.

July 1, 1970 through June 30, 1971

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH

DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities — 1. DCRT - 5

2. PHYSICAL SCIENCES LABORATORY — 3. Dr. G. H. Weiss

I. OBJECTIVES

The Physical Sciences Laboratory is devoted to the study of problems in physics and chemistry that relate to the biological sciences. Several disciplines are represented in the membership of the laboratory. These include applied mathematics, theoretical chemistry, and theoretical physics. Whenever possible the theoretical studies are performed in conjunction with experimental work, either in collaboration with workers in outside units, or by members of the Physical Sciences Laboratory working in other laboratories at NIH. In addition to performing research of its own choosing, members of the Physical Sciences Laboratory provide consultation to other researchers at NIH on different topics in the disciplines represented in the Laboratory. These services are enumerated in the project reports.

II. CURRENT LABORATORY PROGRAMS

Summary

As in earlier years, research in the Physical Sciences Laboratory can be categorized as being in the fields of applied mathematics, theoretical and applied physics, and theoretical and applied chemistry.

1. In applied mathematics, further progress has been made in developing the theory of the ultracentrifuge to elucidate experimental data and to suggest further experimental procedures. In this last year we have concentrated on the effects of hydrostatic pressure on velocity sedimentation experiments. We have suggested a new phenomenological theory of the pressure dependence of the sedimentation coefficient more with results more nearly in agreement with experiments on polymers in compressible solvents. We have further developed a singular perturbation procedure for the approximate solution of equations in ultracentrifugation, allowing the study of general pressure dependence both of sedimentation and of diffusion. The general technique has been successfully applied to a theory of gel pore electrophoresis, and will probably be quite useful in a whole class of similar problems arising from the study of biochemical separation techniques. We have also obtained some exact results on sedimentation in a density gradient.

2. Research has gone forward in the development and application of nonlinear curve fitting routines. The MODELAIDE program has been successfully applied to a large varietyof problems in chemistry and biology, mostly in collaboration with other scientists at NIH, but also with other members of the PSL.

3. In the field of physics we have recently developed a theory of laser light scattering from motile microorganisms. The theory developed so far has been confirmed experimentally by workers at MIT with a correlation-function laser scattering spectrometer. The development both of theory and experiment in this field provides microbiologists with an accurate and rapid measuring device for bacteria and other small organisms. Experiments on chemotaxis and related phenomena are currently being carried out with the techniques and apparatus that have been supported by a contract with the PSL.

4. Work is continuing on the role of van der Waals forces in Biological systems. In the past year we have simplified the Lifshitz theory considerably, allowing us to work on more complicated geometric and dielectric configurations than the theory has so far permitted. A theory of electrostatic interactions between surfaces with ionizable groups has been developed, allowing the calculation of the equilibrium between van der Waals and electrostatic forces. We have also developed the theory of van der Waals forces between media with inhomogeneous dielectric properties, showing large effects due to inhomogeneity. We have also considered the effects of anisotropic dielectric properties, showing that in some cases the forces in different angular configurations provide sufficient torque for effective orientation of the bodies.

5. A considerable effort has been expended on computer assisted NMR studies of proteins. The object of these studies is to develop computer techniques to shed light on the structure and function of enzymes. The use of MODELAIDE on NMR spectra has provided considerable insight into the interaction of the active site histidines in ribonuclease. A recently completed study of the acid denaturation of nuclease by these means has provided definite evidence for a multi-stage, as opposed to a two stage equilibrium during denaturation, a point in contention for a long time between protein biochemists.

6. Work is continuing on the development of equipment and methods to obtain natural abundance C^{13} NMR spectra at 55 MHz using the Fourier transform technique. A Raytheon 704 computer will be interfaced to the spectrometer to perform an on-line Fourier analysis of the data. We have also completed an interactive computer program for the PDP-10 to analyze complex NMR spectra.

7. Calculations have been made on the dielectric relaxation properties of polar synthetic polymers, and a distribution of relaxation times has been found. The calculations allow a prediction of a variety of physical properties of linear molecules with any given sequence of monomer residues.

Serial No. 5.1
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Theory of the Ultracentrifuge

Previous Serial Number: 5.1

Principal Investigator: George H. Weiss, Ph.D.

Other Investigators: None

Cooperating Units: David Yphantis, Ph.D., State University of New York, Buffalo, and
Menachem Dishon, Ph.D., Weizmann Institute, Israel

Man Years

Total:	0.4
Professional:	0.3
Other:	0.1

Project Description:

Objectives:

To determine the effects of various factors such as concentration dependent sedimentation, pressure, density gradients, variations in rotor speed, and polydispersity on current techniques for determining molecular weights. To devise corrections and new techniques of ultracentrifugation which bypass or eliminate these effects.

Methods:

The methods employed involve analytical and numerical solutions of linear and nonlinear partial differential equations.

Major Findings:

An investigation of the effects of pressure on velocity sedimentation experiments has resulted in a new model in which the sedimentation coefficient s_p depends on pressure P as $s_p=s_o(1+\gamma P)^{-1}$ where γ is a constant rather than $s_p=s_o(1-\gamma P)$ as is commonly assumed. Numerical solutions to the Lamm equation have been generated for this pressure dependence, and analytic approximations to the solutions have been generated by a technique developed by Weiss and

Dishon. A major feature of the resulting model is that past the discontinuity the concentration profile is nearly flat for parameter values that arise in practice. This is in accord with an experimental finding of Billick. An investigation is being continued which examines the dependence of diffusion coefficient on pressure. We have also developed a theory of the kinetics of sedimentation in density gradient centrifugation, obtaining exact solutions to the relevant equations in contrast to other investigators who only were able to find partial results. The theory that has been developed allows one to take into account time dependent density gradients as well as preformed gradients.

Significance to Biomedical Research

The ultracentrifuge is very commonly used for the analysis of material of biological interest. In particular density gradient centrifugation is a useful tool for the separation of material such as viruses with slightly differing physical properties. The theory that has been developed will allow better planning of experiments involving density gradients. In addition we have found a simple technique involving gradients that allows one to determine the sedimentation constant rapidly and accurately.

Honors and Awards: None

Publications:

Nossal, R.J. and Weiss, G.H.: Sedimentation in a time varying ultracentrifuge. Analytical Biochemistry, 38, 115-120, 1970.

Dishon, M., Weiss, G.H., and Yphantis, D.A.: Numerical solutions to the Lamm equation VI: Effects of hydrostatic pressure on velocity sedimentation of two component systems, Polymer Science Physics, 8, 2163-2176, 1970.

Dishon, M., Weiss, G.H., and Yphantis, D.A.: A new approach to the effects of pressure on velocity sedimentation experiments, Polymer Science Physics, (to appear).

Dishon, M., Weiss, G.H., and Yphantis, D.A.: Kinetics of sedimentation in a density gradient, Biopolymers (to appear).

Serial No. 5.2
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Theory of the Helix-Random Coil Transformation of Polypeptides in Solution

Previous Serial Number: 5.2

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Robert Jernigan, Ph.D. and Julia Milstein, Ph.D. (NIAMD)

Cooperating Units: Laboratory of Physical Biology, NIAMD

Man Years:

Total:	0.30
Professional:	0.30
Other:	0.00

Project Description:

Objectives:

To gain insight into the mechanism of the helix-random coil transformation and the activity of proteolytic enzymes toward water soluble polypeptides.

Methods:

Examination of an exactly solvable kinetic model of the transformation where both nucleation and chain length effects are included.

Major Findings:

The kinetic model, which correctly explained the NMR and kinetic data on the helix-random coil transformation, correctly predicts the change in the rate of degradation of water soluble polypeptides by proteolytic enzymes as a function of pH.

Significance to Biomedical Research:

The polypeptides are a model chemical system that simulates many of the kinetic features of DNA. It is hoped that a study of polypeptide

kinetics will shed light on the kinetics of DNA coiling and uncoiling as well as other biologically interesting processes.

Honors and Awards: None

Publications: None

Serial No. 5.3
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer Analysis of Nuclear Magnetic Resonance Spectra

Previous Serial Number: 5.3

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Norman Sharpless, Ph.D., (NIAMD); Robert Bradley, Ph.D., (NIAMD); Robert Highet, Ph.D., (NHI); Rolf Johannesen, Ph.D., (NBS).

Man Years:

Total:	1.5
Professional:	1.0
Other:	0.5

Project Description:

Objectives:

To analyze nuclear magnetic resonance spectra of complex molecules in terms of their chemical shifts and spin-spin interaction parameters.

Methods:

An interactive computer program which takes advantage of magnetic symmetry was written to analyze the NMR spectra. A routine for the PDP-10 computer was also written to display the calculated spectra.

Major Findings:

It is possible to analyze very complex spin systems. Analysis of a series of fuzed ring hydrocarbons yielded detailed information on the nature of substituent effects on adjacent carbon atoms. The use of the PDP-10 display in an interactive mode is a very quick and efficient way to perform the analysis.

Significance to Biomedical Research:

This is apparently the most sophisticated program available for the analysis of NMR spectra. The use of a graphic display as an aid in the analysis for NMR spectra is unique.

Honors and Awards: None

Publications:

Johannesen, R. B., Ferretti, J. A., and Harris, R. K.: UEAITR: A New Computer Program for Analysis of NMR Spectra. Analysis of the Proton Spectrum of Triisoprylphosphine. Journal of Magnetic Resonance 3, 84 (1970).

Serial No. 5.4
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Molecular Mechanics

Previous Serial Number: 5.4

Principal Investigator: Robert L. Jernigan, Ph.D.

Other Investigators: None

Man Years:

Total:	0.9
Professional	0.9
Other	0.0

Project Description:

Objectives:

The purpose is to predict dilute solution behavior of macromolecules by means of detailed molecular models. The perturbing influences of external fields on properties as well as the relaxation of these properties upon removal of the fields have been treated.

Methods:

Experimental information on small molecules is combined with a one-dimensional statistical conformational model to permit calculation of properties of large linear molecules. Each bond is permitted to assume several fixed-geometry states. For equilibrium properties, nearest neighbor interactions between these states, as incorporated by matrix methods, suffice to produce agreement between numerous calculated results and experimental data. Series expansions in powers of the external field strength are utilized to estimate these effects. Non-equilibrium properties are treated by assigning rates for transitions between states for each bond. All calculated results for relaxations have been performed by assuming these rates to be independent of the states of neighboring bonds. A more complex and realistic treatment which includes the neighbor dependence of these rates has been formulated but has not yet been completed.

Major Findings:

The effect of electric fields on some polar synthetic polymers was calculated and found to be adequately represented by the first few terms in a series expansion. The dielectric relaxation of these molecules was also investigated. Because of the independence of conformational transitions, no dominant long relaxation time was found. Instead a wide distribution of relaxation times was observed.

Significance to Biomedical Research:

These calculations allow a prediction of a variety of physical properties for linear molecules with any given sequence of monomer residues, whether in biological or synthetic macromolecules. Performing such calculations in advance of laboratory experiments can give the experimental scientist a guide to indicate the value of particular experiments.

Honors and Awards: None

Publications:

Jernigan, R. L.: Internal Relaxation in Short Chains Bearing Terminal Polar Groups. Polymer Preprints Vol. 12, No. 1, March 1971.

Serial No. 5.5
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Biophysical Analysis

Previous Serial Number: 5.5

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: Michael C. Mackey, Ph.D. and James E. Kiefer

Cooperating Units: Harold Lecar, Ph.D., Biophysical Laboratory, NINDB

Man Years:

Total:	.60
Professional	.60
Other	.00

Objectives:

To supply theoretical foundations and experimental models for various observations which arise in physiology and biophysical chemistry. The following problems have received particular attention during the past year.

1) Models for macromolecular rate processes: The objective is to provide mathematical foundations for various kinetic phenomena arising in studies of macromolecular biochemistry.

2) Phenomena occuring in phospholipid bilayer membranes: The objective is to develop experiments and theories relating conductivity measurements and transport data to the molecular processes which give rise to excitability in biomembranes.

3) Threshold fluctuations in nerves: The objectives are to relate the fluctuations in firing thresholds of nerves to the chemical and physical processes underlying excitation and to analyze existing data in order to test theories concerning mechanisms of transport of ions across nerve membranes.

Methods:

1) Kinetic equations have been derived and investigated for a number of possible association mechanisms.

2) Laboratory techniques of microbiology and electrophysiology are combined with techniques of mathematical physics.

3) Existing neurophysiological equations (the Hodgkin-Huxley equations) have been modified to include fluctuating forces. Methods of non-linear mechanics and statistical physics have been applied to analyze the equations and digital computers are used to determine phase trajectories.

Major Findings:

1) Solutions have been obtained to equations which describe the time dependent attachment and detachment of antibodies to viruses.

2) Data have been taken regarding the voltage dependent ionic conductivity of oxidized cholesterol films.

3) Comparison with data indicate that threshold fluctuations are due to the Na^+ conductivity processes.

Significance to Biomedical Research:

1) Kinetic models: The theoretical studies describe systems which are analogs of various physiological phenomena.

2) Reconstituted phospholipid membranes: The membrane components can be controlled by the experimentalist, thus providing an opportunity to devise procedures and test theories applicable to natural biological membranes.

3) Threshold fluctuations: It is believed that the study will facilitate better understanding of the physical processes underlying the excitation and propagation of nervous impulse.

Publications:

Kiefer, J. E. and Nossal, R.: Solution of Equations Describing a Model of Attachment and Detachment of Antibodies to Viruses. Mathematical Biosciences (to appear).

Serial No. 5.6
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Carbon-13 Nuclear Magnetic Resonance Studies of Peptides and Proteins, Including the Use of Selective C^{13} Enrichment

Previous Serial Number: None

Principal Investigator: Jack S. Cohen, Ph.D.

Other Investigators: I. Chaiken, Ph.D. and M. Freedman Ph.D.

Cooperating Units: Laboratory of Chemical Biology, NIAMD

Man Years:

Total:	0.2
Professional:	0.2
Other:	0.0

Project Description:

Objectives:

1. To investigate the applicability of the recently developed carbon-13 Fourier transform NMR methods to macromolecules.

2. To obtain information regarding the structure and function of peptides and proteins.

Methods:

1. Synthesis of peptides containing selected C^{13} enriched amino acids.

2. Use of C^{13} Fourier Transform NMR to study these resonances in the peptides and on the formation of active enzyme complexes.

3. To study the pH and other dependences of C^{13} resonances of amino acids and peptides.

Major Findings:

1. A peptide corresponding to the 1-15 sequence of the amino-terminal region of ribonuclease was synthesized containing 15% C^{13} enriched

phenylalanine at position 8. Its C^{13} resonances could be readily seen above the background of those due to the natural abundance C^{13} (1%) of the other 14 amino acids.

2. Comparison of spectra of the 1-13, 1-15, and 1-20 peptides with each other and with amino acid spectra, indicate

a) that some resonances may be definitively assigned,
b) that some resonances of the same kind of residue must be non-equivalent or shifted, even in a peptide normally supposed to have no tertiary structure.

3. The pH-dependence of C^{13} resonances of several amino acids and peptides is somewhat surprising. Little effect of the carboxyl titration was observed, but fairly large effects of the amino terminus titration. This is being investigated for possible applications.

Publications:

Cohen, J.S., Horsley, W. and Sternlicht, H.: Isolation of Carbon-13 Enriched Amino Acids. Biochem. Biophys. Acta., 222, 521 (1970).

Freedman, M., Cohen, J.S. and Chaiken, I.: Carbon-13 Fourier Transform Nuclear Magnetic Resonance Studies of Peptides. Biochem. Biophys. Res. Comm., in press.

Serial No. 5.7
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Reports
July 1, 1970 through June 30, 1971

Project Title: Excitation and Transport Properties of Fluids; Laser Scattering

Previous Serial Number: 5.7

Principal Investigator: Ralph J. Nossal, Ph.D.

Other Investigators: None

Cooperating Units: John Utting, Ph.D., LBP, NIAMD

Man Years:

Total:	.50
Professional:	.50
Other:	.00

Project Description:

Objectives:

To provide basic knowledge concerning the excitation properties and transport properties of both simple fluids and complex solutions containing biological macromolecules.

To develop a laser light scattering spectrometer to measure hydrodynamic coefficients of biological macromolecules, the rate constants of biomolecular reactions, and the swimming speed distributions of motile microorganisms.

Methods:

Theoretical techniques of mathematical physics and statistical mechanics are employed in order to develop new physical theories. Theoretical studies are performed in support of new experiments, particularly those involving the scattering of laser light from biological materials. A correlation-function laser scattering spectrometer is being developed under contract.

Major Findings:

Techniques have been developed for using a laser light scattering correlation spectrometer for the purpose of measuring the diffusion coefficients of biological molecules having a wide range of sizes (even as small as MW ~ 1000). The theoretical basis of light scattering from motile microorganisms has been investigated, and spectra associated with various distributions of swimming speeds have been calculated. Experiments have been performed which establish the feasibility of using laser light scattering spectroscopy as a new quantitative assay for bacterial motility.

Significance to Biomedical Research:

Almost all biological phenomena occur in a fluid environment. A number of fundamental questions concerning the physical behavior of fluids yet remain unanswered. Their elucidation will ultimately facilitate better understanding of the functions and properties of biological systems. The development of the laser correlation spectrometer may enable rapid and precise measurement of various physical parameters characterizing systems of biological molecules.

Honors and Awards: None

Publications:

Nossal, R.: Spectral Analysis of Laser Light Scattered from Motile Microorganisms. Biophysical Journal (to appear).

Serial No. 5.8
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Intermolecular Forces in Biological Structures

Previous Serial Number: 5.8

Principal Investigators: V. Adrian Parsegian, Ph.D., George H. Weiss, Ph.D., and David Gingell, Ph.D.

Other Investigators: None

Cooperating Units: None

Man Years:

Total:	2.2
Professional	2.0
Other	0.2

Project Description:

Objectives:

To identify and learn to calculate those intermolecular forces governing the structure of matter at the biological level. These have been Coulombic (electrostatic) interactions between charged species and van der Waals (electrodynamic) forces.

Methods:

Classical and quantum-mechanical treatment of electromagnetic and statistical-mechanical behavior as well as molecular models of specific structural interactions.

Major Findings:

We have developed the theory of electrostatic interactions between surfaces bearing ionizable groups, allowing a calculation of the equilibrium achieved by balancing electrostatic and van der Waals forces. We have also developed theories applicable to the interaction between media with inhomogeneous dielectric properties and between media with anisotropic dielectric properties.

Significance to Biomedical Research:

It is known that van der Waals forces are important factors in determining the interaction of biological surfaces. Earlier calculations of these forces used the assumption of pairwise additivity, which can lead to grossly incorrect results. The calculations that have been carried out under this project correctly evaluate many-body contributions.

Honors and Awards: None

Publications:

Ninham, B. W. and Parsegian, V. A.: Interactions in Multilayer Systems. Journal of Chemical Physics, 53, 3398-3402, 1970.

Ninham, B. W., Parsegian, V. A., and Weiss, G. H.: On the Macroscopic Theory of Temperature Dependent van der Waals Forces. Journal of Statistical Physics, 2, 323-328, 1970.

Parsegian, V. A., and Ninham, B. W.: Temperature-Dependent van der Waals Forces. Biophysical Journal, 10, 664-674, 1970.

Ninham, B. W., and Parsegian, V. A.: Van der Waals Forces: Special Characteristics in Lipid-Water Systems and a General Method of Calculation Based on the Lifshitz Theory. Biophysical Journal, 10, 646-663, 1970.

Ninham, B. W., and Parsegian, V. W.: Electrostatic potential Between Surfaces Bearing Ionizable Groups in Ionic Equilibrium with Physiologic Saline Solution. Journal of Theoretical Biology (to appear).

Parsegian, V. A. and Ninham, B. W.: Toward the Correct Calculation of van der Waals forces Between Lyophobic Colloids in Aqueous Solution. Journal of Colloid and Interface Science (to appear).

Serial No. 5.9
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Consulting Services

Previous Serial Number: 5.9

Principal Investigators: Richard I. Shrager, George H. Weiss, Ph.D.

Other Investigators: Mildred McNeel

Cooperating Units: John Folk, M.D., Ron Chung, M.D., Laboratory of Biochemistry, NIDR; Leonard Kohn, M.D., Laboratory of Physical Biology, NIAMD; Elemer Mihaly, Ph.D., Laboratory of Biochemistry, NHLI; John Minner, M.D., Arthur J. Blume, Laboratory of Biochemical Genetics, NHLI; John R. Wunderlich, M.D., Immunology Branch, NCI; Steven R. Levisohn, M.D., Laboratory of Biochemistry, NCI; Marc Lewis, Ph.D., Laboratory of Biochemistry, NEI; William Caveness, M.D., Laboratory of Experimental Neurology, NINDS.

Man Years:

Total:	1.2
Professional	1.0
Other	0.2

Project Description:

Objective:

To provide consulting services in applied mathematics, biometry, theoretical chemistry, theoretical physics, and various aspects of numerical analysis to workers primarily in experimental fields.

Methods:

A considerable amount of consulting is done in the application of nonlinear curve fitting techniques applied to the results of biological and biochemical experiments. A smaller amount of consulting is given on purely statistical and epidemiological problems.

Major Findings:

The MODELAIDE program and modifications thereof has been used in the elucidation of many types of biological and biochemical experiments. Some specific applications have been:

1. A study of the time course of the digestion of fibrogen.

2. Enzyme regulation by steroids in mammalian cells.

3. Immune destruction of tumor cells by spleen cells.

4. Enzyme production in growing neuroblastoma cells.

5. Ultracentrifuge studies of hemoglobin binding and polymer formation.

Data on head injured veterans of the Korean war has been examined in order to find prognostic indicators of subsequent fits. We have shown that the type of injury and related factors are not too useful in predicting the course of the fit behavior. It has also been found that the time of onset of initial fit and the initial pattern of fits is a useful prognostic tool.

Significance to Biomedical Research:

Numerical methods have become a useful adjunct to research in the biomedical sciences. Curve fitting programs allow one to test and discriminate between different models.

Honors and Awards: None

Publications:

Chung, S. I., Shrager, R. I., and Folk, J. E.: Mechanism of Action of Guinea Pig Liver Transglutaminase, VII. Chemical and Sterochemical Aspects of Substrate Binding and Catalysis. Journal of Biological Chemistry, 245, 6424-6435, 1970.

Serial No. 5.10
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Fundamental Studies

Previous Serial Number: 5.10

Principal Investigators: George H. Weiss, Ph.D., Ralph J. Nossal, Ph.D., Andrew G. De Rocco, Ph.D., Richard I. Shrager

Other Investigators: James E. Kiefer, Mildred McNeel

Cooperating Units: Milton Sobel, Ph.D. University of Minnesota, Leonard Rodbard,M.D., NICHD

Man Years:

Total:	1.5
Professional	1.3
Other	0.2

Project Description:

Objective:

This project presently encompasses several lines of investigation, not all of which relate to biomedical problems. These are:

1. The study of statistical procedures for the design and evaluation of clinical trials.

2. A study of perturbation procedures in the mathematical theory of epidemics.

3. The study of molecular forces between pairs of polyatomic systems, with emphasis on the additivity or nonadditivity of such forces.

4. The study of gel pore electrophoresis for the purpose of designing gel gradients for optimal separation of proteins.

Methods:

A variety of methods are necessary for the problems mentioned above. These include the theory of stochastic processes, classical analysis,

and methods of mathematical physics.

Major Findings:

We have continued work on a project to design clinical trials to distinguish between two drugs with a given level of confidence in such a way as to minimize the number of patients on the poorer drug. Along these lines we have developed a fixed sample design which performs as well as the unlimited sample designs. We have also devised a two stage adaptive design which will take advantage of some earlier work, to further reduce the number of patients on the poorer drug. On the second project we have developed a model for which it can be proved that the deterministic result gives a result that differs from that given by the stochastic theory. However, we can show the times at which differences become apparent tend to infinity with the numbers in the population.

A detailed model for the phase transition between anisotropic and isotropic phases for long, semi-flexible molecules was completed. This study is of interest in connection with liquid crystals and the role such arrays may play in membrane phenomena. The model developed has the advantage that it is the first to account for the transition in the correct density regime and with suitable transitional parameters.

The role of clathrate structures in general anesthesia was investigated from the viewpoint of a statistical-thermodynamic theory of the formation of clathrates in vivo. Particular attention was paid to the role which can be played by protein sidechains, by ions and by the neurotransmitters themselves. The model developed for anesthesia permits a calculation of the pressures required to maintain anesthesia (third plane, third stage) and the calculated values agree within experiment error to the observed values.

A further study of the interaction potential between pairs of large polyatomic molecules was made. An extension of earlier results leads to the observation that large, symmetric systems behave in the main as if their interaction arises from the interactions of the constituent atoms in pairs.

A calculation of the perturbation introduced into the forth virial coefficient for a hard sphere system when a soft long-range attractive tail is added to the potential was done for the case that the tail is linear with distance (a triangle well). The results were shown to be better than those for a square-well and nearly as good as those for a Lennard-Jones potential. Because the triangle-well is much simpler to handle mathematically, there is some hope that it may be useful to study in a fairly realistic way, more complicated problems than are now possible for a Lennard-Jones potential.

On the fourth project we have developed a theory of gel pore electrophoresis in which the ratio of free mobility to mobility in the gel is equal to the ratio of free diffusion constant to that in the gel. The results are exact for a linear gel gradient. An approximate theory has been developed which applies when diffusion is a small effect relative to electrophoretic mobility. The theory can be used for general dependence of mobility and

diffusion on gel concentration.

Significance to Biomedical Research:

The work on clinical trials is obviously and directly related to many types of comparative clinical trials. Gel pore electrophoresis and similar biochemical separation techniques are widely used to characerize molecules of biological interest. The theory developed will be used to design gradients for optimal separation of proteins.

Honors and Awards: None

Publications:

De Rocco, A. G., and York, E. D.: Nonadditivity in the potential for pairs of polyatomic molecules. Journal of Chemical Physics, 53, 764-767, 1970.

Weiss, G. H., and Gillis, J.: On the expected number of distinct sites visited by a random walk with an infinite variance. Journal of Mathematical Physics, 11, 1308-1313, 1970.

Weiss, G. H., and Sobel, M.: Inverse sampling and other selection procedures for tournaments with 2 or 3 players. Proceedings of the Conference on Non-Parametric Statistics ed. M. Puri (Cambridge University Press, Cambridge) 515-543, 1970.

Weiss, G. H., and Blumenfeld, D.: On queue splitting to reduce waiting times. Transportation Research, 4, 141-144, 1970.

Weiss, G. H., and Blumenfeld, D.: On the robustness of certain assumptions in the intersection delay problem. Transportation Research, 4, 125-139, 1970.

Weiss, G. H.: On the noise generated by a stream of vehicles. Transportation Research 4, 229-234, 1970.

Weiss, G. H., and Blumenfeld, D.: Routing in a circular city with two ring roads. Transportation Research, 4, 235-242, 1970.

Weiss, G. H. and Sobel, M.: Play-the-winner sampling for selecting the better of two binomial populations. Biometrika, 57, 357-365, 1970.

Weiss, G. H., and Sobel, M.: A comparison of play the winner and vector at a time sampling for selecting the better of two binomial populations with restricted parameter values. Trabajos de Estadisticas y de Investigacion Operativa (to appear).

Weiss, G. H., and Dishon, M.: On the asymptotic behavior of the stochastic and deterministic models of an epidemic. Mathematical Biosciences (to appear).

Shrager, R. I.: Nonlinear regression with linear constraints: An extension of the magnified diagonal method. Journal of the Association for Computing Machinery, 17, 446-452, 1970.

Serial No. 5.11
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-MS
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer-Assisted Nuclear Magnetic Resonance (NMR) Studies of Proteins

Previous Serial Number: 5.11

Principal Investigator: Jack S. Cohen, Ph.D.

Other Investigators: A. Schechter, M.D., M. McNeel, R.I. Shrager, H. Epstein, M.D., D. Sachs, M.D., S. Heller, Ph.D., M. Freedman, Ph.D. and H. Pollard, M.D.

Cooperating Units: Laboratory of Chemical Biology, NIAMD

Man Years:

Total:	.75
Professional:	.75
Other:	.00

Project Description:

Objectives:

1. To obtain detailed information on the structure and function of enzymes.

2. To extend the applicability of high resolution NMR to proteins.

Methods:

The region of the NMR spectrum containing the histidine imidazole C2-H resonances of a number of proteins has been subjected to detailed analysis. The 220 MHz time-averaged spectra were digitized and fitted with a series of Lorentzian curves using the MODELAIDE least-squares fitting program with the on-line IBM 2250 display unit. The chemical shifts of the imidazole resonances as a function of pH have been fitted to several mechanistic models.

Major Findings:

1. The relative areas of peaks in the spectra of staphylococcal nuclease were measured thus aiding in the identification of the four imidazole resonances.

Major Findings:

1. The relative areas of peaks in the spectra of staphylococcal nuclease were measured thus aiding in the identification of the four imidazole resonances.

2. The acid denaturation process was followed by measuring the areas of the peaks. The results indicated that the renaturation process was reversible and that it occurs via multiple equilibria with different nucleation sites.

3. The fitting of the titration data to single normal equilibria enables the precise determination of the individual histidine microscopic dissociation constants (pK) and allows an objective criterion for deciding the true continuity when curves cross. The finds for S. nuclease do not indicate a conformational equilibrium involving a histidine residue, contrary to the interpretation by Markley and Jardetzky of their (100 MHz) NMR data.

4. Titration curves of the proton resonances of imidazole in histidine, its blocked derivatives and small peptides show the effects of the adjacent titrating carboxyl and amino groups.

5. Extensive data on the titration curves of the two active site histidines (at positions 12 and 119) in ribonuclease has been accumulated. An equation describing the titration of two interacting titrating groups in terms of chemical shift has been derived. When applied to the asymmetric curves of active site histidines this indicates that, contrary to a commonly held view, the two histidines are not interacting with each other, but are both interacting with a carboxyl group, probably that of aspartic acid-121.

6. A study of bovine and human carbonic anhydrases has indicated that no histidine group is directly involved in their enzymatic mechanism. However, a conformational change on binding inhibitor, resulting in the observation of an extra histidine titration-curve with a pK of 6.04, occurs in HCA.

7. Preliminary studies with myoglobin indicate several well-resolved titrating histidine resonances.

Significance to Biomedical Research

1. A general and objective method for the analysis of resolved resonances in protein NMR spectra has been developed.

2. A recently completed study of the acid denaturation of nuclease provides definite evidence for a multi-state, as opposed to a two-state, equilibrium during denaturation, a point in contention for a long time between protein biochemists.

3. An analysis of the titration data for pancreatic ribonuclease is under way. This approach provides in principle a means of distinguishing between published mechanisms of its activity.

Honors and Awards: None

Publications:

Cohen, J. S., Shrager, R. I., McNeel, M., and Schechter, A. N.: On-line Computer-Assisted Analysis of 220 MHz NMR Data of Protein Imidazole Resonances. Biochem. Biophys. Res. Comm. 40, 144, 1970.

Cohen, J. S., Shrager, R. I., McNeel, M., and Schechter, A. N.: Proton Magnetic Resonance Studies at 220 MHz of the Histidine Residues of Staphylococcal Nuclease. Nature, 228, 642, 1970.

Cohen, J. S.: Proton Magnetic Resonance Studies of Oligopeptides Containing Aromatic Residues. Biochim. Biophys. Acta, in press.

Serial No. 5.12
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: High-Resolution Carbon-13 Fourier Transform NMR Spectroscopy at 55 MHz

Previous Serial Number: 5.12

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Edwin D. Becker, Ph.D., LPB:NIAMD, Perry Plexico, CSL; Arthur Schulz, CSL; Norman LeClair, CSL; and Jack Kane, DRS

Cooperating Units: Laboratory of Physical Biology, NIAMD and Bioelectrical Engineering, DRS

Man Years:

Total:	3.0
Professional	3.0
Other	0.0

Project Description:

Objectives:

To develop equipment and methods to obtain natural abundance carbon-13 NMR spectra at 55 MHz (ca. 51 kgauss) using the Fourier transform technique.

Methods:

Equipment is being constructed, preliminary tests are being performed, programs are being written, and a Raytheon 704 computer is to be interfaced to the spectrometer to permit redording the free induction decay in a pulse NMR experiment and then taking the Fourier transform after sufficient signal enhancement.

Major Findings:

Preliminary experiments indicate that the Fourier transform technique will provide an efficient means for studying carbon-13 spectra in natural abundance.

Significance to Biomedical Research:

Carbon-13 NMR provides a sensitive technique to correlate molecular changes with biological activity such as changes in conformation and active binding sites in enzymes.

Honors and Awards: None

Publications: None

Serial No. 5.13
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Rate Studies by Nuclear Magnetic Resonance Spectroscopy

Previous Serial Number: 5.13

Principal Investigator: James A. Ferretti, Ph.D.

Other Investigators: Stephen R. Heller, Ph.D., DMB; Richard I. Shrager and Mildred McNeel

Cooperating Units: None

Man Years:

Total:	0.50
Professional	0.50
Other:	0.00

Project Description:

Objectives:

By a very detailed error analysis, to determine the limits of the NMR technique in rate studies.

Methods:

Use of the programs written for chemical exchange by NMR in conjunction with MODELAIDE and the IBM 2250 graphic display.

Major Findings:

The NMR technique for rate studies is subject to a number of random and non-random errors, and is not as accurate as was previously thought.

Significance to Computer Technology:

These results allow one to reasonably estimate the errors associated with the analysis of NMR data on rate studies.

Honors and Awards: None

Publications: None

Serial No. 5.14
1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Gas-Liquid Chromatography-Mass Spectrometry of Isotopically Labeled Compounds

Previous Serial Number: None

Principal Investigator: Jack S. Cohen, Ph.D.

Other Investigators: W. VandenHeuval, Ph.D.

Cooperating Units: Merck Institute, Rahway, New Jersey

Man Years:

Total:	0.2
Professional	0.2
Other:	0

Project Description:

Objectives:

1. To determine isotopic enrichments of substances from very small quantities containing mixtures.

2. To obtain structural information on the fragment ions formed during mass spectrometry.

Methods:

Isotopically enriched (C^{13} ^{18}O) compounds were prepared and subjected to combined GLC-MS analysis

Major Findings:

1. A general equation was derived for the calculation of isotopic content from two peaks in the mass spectrum of a fragment ion.

2. Variations in the C^{13} content for the side chain residues but not the backbone, of the enriched amino acids isolated from algae grown on enriched carbon dioxide indicated a net C^{13}biosynthetic isotope effect.

3. Deuterated amino acids could be separated on GLC but C^{13} enriched amino acids could not.

4. Some mass spectral fragment ions have been identified by the use of isotopic enrichment.

Significance to Biomedical Research:

1. It is now possible using GLC-MS to carry out metabolism studies with small quantities of enriched isotopic (C^{13}) material and to determine the resultant enrichment at selected positions in the product molecule.

2. The observation of a macroscopic C^{13} biosynthetic isotopic effect confirms previous tentative conclusions and enables individual enzymatic reactions to be investigated for the mechanism at individual carbon atoms.

Honors and Awards: None

Publications:

VandenHeuval, W. J. A. and Cohen, J. S., "GLC-MS of Carbon-13 Enriched Amino Acids as Trimethylsilyl Derivatives", Biochim. Biophys. Acta., 208, 251, (1970).

VandenHeuval, W. J. A., Smith, J. L., Putter, I. and Cohen, J. S., "GLC-MS of of Deuterium-Containing Amino Acids as Trimethylsilyl Derivatives", J. Chromatog., 50, 405 (1970).

VandenHeuval, W. J. A., Smith, J. L., and Cohen, J. S., "GLC-MS of Carbon-13 and Deuterated Amino Acids as Trimethylsilyl Derivatives", J. Chromatog. Sci., 8, 567 (1970).

Cohen, J. S. and Putter, I., "The Isolation of Deuterated Amino Acids", Biochim. Biophys. Acta., 222, 515 (1970).

VandenHeuval, W. J. A., Arison, B., and Cohen, J. S., "Combined GLC-MS Study of the Bromine Oxidation of Synkavit", J. Chromatog., submitted.

Serial No. 5.15

1. Physical Sciences Laboratory
2. Not Applicable
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Models for the Transport Properties of Membranes

Previous Serial Number: 5.15

Principal Investigator: Michael C. Mackey, Ph.D.

Other Investigators: Mildred McNeel

Cooperating Units: None

Man Years:

Total:	0.8
Professional	0.8
Other	0.0

Project Description:

Objectives:

To develop a theory of the electrical properties of membranes based on the ideas and techniques of kinetic theory.

Methods:

The methods common to statistical mechanics and the kinetic theory of gases have been used.

Major Findings:

The kinetic properties of the membrane have been shown to depend on the force law between the ion and membrane. Detailed properties predicted by this model, such as chord conductance, have been shown to be in agreement with some experimentally observed features of the electrical properties of membranes.

Significance to Biomedical Research:

The transport of ions across membranes is a major biological phenomenon, and one that has challenged theorists for many years. The present approach has promise as a basic physical model rather than a phenomonological one.

Honors and Awards: None

Publications:

Mackey, Michael C.: Kinetic Theory Model for Ion Movement Through Biological Membranes. I. Field Dependent Conductance in the Presence of Solution Symmetry. Biophysical Journal V.11, No. 1, 75-91, 1970.

Mackey, Michael C.: Kinetic Theory Model for Ion Movement Through Biological Membranes. II.Interionic Selectivity. Biophysical Journal V. 11, No. 1, 91-97, 1970.

July 1, 1970 through June 30, 1971

PUBLIC HEALTH SERVICE - NATIONAL INSTITUTES OF HEALTH
DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY

Summary of Branch Activities

1. DCRT 6
Serial Number

2. HEURISTICS LABORATORY

3. James R. Slagle
Chief

I. OVERVIEW

The founding of the Heuristics Laboratory was based on the proposition that computers can be programmed to exhibit some intelligent behavior and all the techniques developed can be applied, directly or indirectly, to solve biomedical problems. For example, a graph searching procedure can be used to match chromosomes, techniques used in game playing programs can be used to optimize diagnosis and some results in formal logic can be used to solve medical resource allocation problems.

Perhaps due to close cooperation among researchers in this Laboratory, we have been witnessing an unusual and encouraging phenomenon. That is, we have successfully combined several seemingly unrelated fields. For example, learning (frequently used in game playing) and the concept of GPS (General Problem Solver) are being merged into mechanical theorem proving and formal logic is now used in designing pattern recognizers.

This Laboratory is presently involved in six different projects. They are: (1) deductive question-answering and problem solving, (2) storage and retrieval algorithms, (3) pattern recognition, (4) chemical notation and on-line information systems, (5) on-line interactive graphic curve fitting systems and (6) computer system support. We shall describe the progress made in each of these projects in the next section.

II. SUMMARY OF PROGRESS MADE DURING THE PAST YEAR

(1) Deductive Question-Answering and Problem Solving

Our goal is to develop general purpose question-answering systems and we place our emphasis upon answering difficult questions. It is universally recognized that some deductive power is absolutely necessary for these kinds of question-answering systems. Therefore, we transform the original problem into a theorem proving problem, where facts are stored as axioms and the question is stored in the form of a theorem to be proved. We have continued to make progress in inventing new theorem proving procedures. Six new strategies have been invented. Their theoretical properties have all been studied and three of them have been implemented by LISP programs. Another project used learning to improve a theorem prover's performance.

Many problems can be represented by graphs. Research in this area has yielded two interesting algorithms. One algorithm is concerned with a special kind of graph and the other is an optimal graph searching procedure.

To date, only two-valued logic has been used in problem-solving systems. We have investigated the possibility of using multi-valued logic and some theoretical results have been obtained.

(2) Storage and Retrieval Algorithms

Storage and retrieval schemes are extremely important for many projects inside and outside of our Laboratory. During this past year, a balanced tree storage and retrieval algorithm has been developed.

(3) Pattern Recognition

Research in this area has continued to yield useful results. Combining symbolic logic with linear inequalities, we can now investigate the pattern classification capability of some multi-layered threshold logic machines. Since biomedical problems often involve large amounts of data, the reduction of data is a very important problem. The introduction of Fisher's criterion into pattern recognition seems to yield very encouraging results.

(4) Chemical Notation and On-line Information Systems

There are two basic WLN (Wiswesser Line Notation) programs being developed for the PDP-10 computer. One accepts as input the atom-by-atom connection table of a compound and generates the corresponding WLN. The other basic program converts an input string of WLN to its corresponding connection table. A program then generates a display of the structural diagram from the connection table. The WLN generation program can correctly encode more than 90% of a set of published compounds tested for carcinogenesis.

(5) On-line Interactive Graphic Curve Fitting System

The MODELAIDE package is a very useful tool for scientists who have a set of data and want to construct a mathematical model to fit these data. At present, we have undertaken to rebuild the MODELAIDE system on the PDP-10. This work is still in progress. The completion of this project will not only provide an improved and upgraded MODELAIDE system, but also will allow DCRT to dispense with the IBM 2250 display.

(6) The System Support Project

Some members of this Laboratory have contributed to the upgrading of the PDP-10 and IBM-360 service facilities. Many software packages developed at Stanford University have been implemented here. A link between the IBM-360 and the PDP-10 has been designed and an improvement of the present WYLBUR system has been studied.

Serial No. DCRT 6.1

1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Automatic Question-Answering and Problem Solving

Previous Serial Number: Same

Principle Investigator: James R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D., John K. Dixon, Ph.D., Louis Hodes, Ph.D., Thomas L. Jones, Ph.D., Richard C. T. Lee, Ph.D., Lewis M. Norton, Ph.D.

Cooperating Units: None

Man Years:

Total:	5.95
Professional:	4.95
Other:	1

Project Description:

The object of this research is to develop efficient procedures for computers to answer questions and solve problems. Our approach is to transform the original question answering problem into a theorem proving problem. Thus, the success or failure of our question answering program depends heavily upon the efficiency of theorem proving programs developed in this Laboratory.

Progress Made During the Past Year:

(1) It has been proved that renamable resolution plus renamable paramodulation is complete for functionally reflexive systems. With this procedure, we cannot only dispense with the troublesome axioms for equality, but also reduce the number of intermediate deductions we have to generate in the course of a proof.

(2) A new resolution, called variable-constrained resolution, has been proposed. Experimental results show that this newly invented inference rule is much more efficient than ordinary resolution.

(3) In the usual formulation of problem-solving systems, there are a set of objects and a set of operators. Starting with an initial object, one tries to find a sequence of operators which will transform the initial object

into a desired object. With this formulation, the problem of finding a linear refutation with a top clause can be viewed as a problem-solving system. A paper published by Chang and Slagle thus provides the theoretical basis for a heuristic program to find linear refutations for theorems in first order predicate calculus with equality.

(4) A theorem proving program employing both paramodulation and resolution has been written. It is augmented by numerous heuristics without which many theorems could not be proved.

(5) To make programs using the resolution principle more efficient, an approach has been proposed to take advantage of the structure of special theories, for example, the theories of equality (paramodulation), partial ordering and sets. This new idea, in the case of partial ordering, has been implemented and found effective.

(6) A new inference rule, called Z-resolution, has been proposed. Z-reso- is a method of solving problems by means of compiled axioms. Experimental results have indicated the significance of this inference rule.

(7) A method for eliminating quantifiers in the elementary theory of addition has been programmed and it is useful in solving problems related to linear programming. This approach is, in general, more efficient than the Herbrand approach.

(8) A learning capability has been developed and implemented for use with a theorem proving program. Experimental results show that the theorem proving efficiency can be improved by learning.

(9) All the existing problem solving systems use two-valued logic. This means that they cannot handle information involving uncertainty. The possibility of using fuzzy logic has been studied. Many theorems relating fuzzy logic and two-valued logic have been proved.

(10) Since theorem proving procedures often involve graph searching, a parallel effort has been made to study efficient graph searching procedures. It has been found that a special class of graphs can be reduced to trees and many efficient tree searching techniques can be applied. In addition, an optimal (in the sense that a minimum number of nodes are generated) graph searching procedure has been found.

(11) The INSIM (infant simulation) program which was started by T. Jones at MIT has now been converted to the NIH machine. New features have been added into the program.

(12) It has been found that the personnel resource allocation problem can be formulated as a problem of finding prime implicants. A new algorithm has been proposed to generate prime implicants. Using this algorithm, the selection problem can be solved by a branch-and-bound approach which reduces the amount of calculations necessary for finding an optimal solution.

Proposed Course:

Research in this area will continue to emphasize development and testing of new techniques.

Publications:

1. Chang, C.L.: The Unit Proof and the Input Proof in Theorem Proving. JACM, October 1970, 698-707.

2. Chang, C.L.: Renamable Paramodulation for Automatic Theorem Proving. Artificial Intelligence, Vol. 1, 1970, 247-256.

3. Chang, C.L. and Slagle, J.R.: Completeness of Linear Refutation for Theories with Equalities. JACM, Vol.18, No.1, January 1971, 126-136.

4. Chang, C.L. and Slagle, J.R.: An Admissible and Optimal Algorithm for Searching AND/OR Graphs. (Accepted by Artificial Intelligence).

5. Hodes, L.: Solving Problems by Formula Manipulation in Logic and Linear Inequalities. Proceedings of the Second International Joint Artificial Intelligence Conference, London, 1971.

6. Lee, R.C.T.: An Algorithm to Generate Prime Implicants and Its Application to the Selection Problem. (Accepted by Information Science).

7. Lee, R.C.T.: Application of Information Theory to the Variable Selection Problem. (Accepted by Mathematical Biosciences).

8. Lee, R.C.T.: Fuzzy Logic and Resolution Principle. Proceedings of the Second International Joint Artificial Intelligence Conference, London, 1971.

9. Simon, R. and Lee, R.C.T.: On Efficient Ways to Find Optimal Solutions to AND/OR Series-parallel Graphs. (Accepted by JACM).

10. Slagle, J.R.: Artificial Intelligence: The Heuristic Programming Approach. McGraw-Hill, New York, 1971.

11. Slagle, J.R. and Farrell, C.: Experiments in Automatic Learning for a Multipurpose Heuristic Program. Comm. of ACM, February 1971, 91-99.

12. Slagle, J.R.: Automatic Theorem Proving with Built-in Theories Including Equality, Partial Ordering and Sets. (Accepted by JACM).

13. Slagle, J.R.: Interpolation Theorems for Resolution in Lower Predicate Calculus. JACM, Vol.17, July 1970, 535-542.

14. Slagle, J.R.: Heuristic Search Programs. In Theoretical Approach to Non-numerical Problem Solving. (M.D. Mesarovic and R.B.Banerji, Eds.), Springer-Verlag, Berlin, 1970, 246-273.

15. Slagle, J.R. and Koniver, D.: Finding Resolution Proofs and Using Duplicate Goals in AND/OR Trees. (Accepted by Information Science).

16. Chang, C.L.: Theorem Proving with Variable-Constrained Resolution. (Accepted by Information Science).

Serial No. DCRT 6.2

1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Storage & Retrieval Studies

Previous Serial Number: None

Principal Investigator: Gary D. Knott

Other Investigators: None

Cooperating Units: None

Man Years

Total:	0.33
Professional:	0.28
Other:	0.05

Project Description:

The objective of this project is to catalog and analyze storage and retrieval algorithms.

Progress Made During the Past Year:

A balanced tree storage and retrieval scheme which places items to be stored at the nodes of a binary tree has been developed. In particular, a good deletion process has been discovered and described.

Proposed Course:

The study of various algorithms is continuing. A major effort is being made in discovering and describing so-called hashing functions used in computing addresses.

Publications:

1. Knott, G.: A Balanced Tree Storage and Retrieval Algorithm. Proceedings of the 1971 ACM Symposium on Information Storage and Retrieval, 175-196.

Serial No. DCRT 6.3

1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Automatic Pattern Recognition

Previous Serial Number: Same

Principal Investigator: J.R. Slagle, Ph.D.

Other Investigators: Chin-Liang Chang, Ph.D., Louis Hodes, Ph.D., Thomas L. Jones, Ph.D., Richard C. T. Lee, Ph.D.

Cooperating Units: None

Man Years

Total:	2.06
Professional:	1.81
Other:	0.25

Project Description:

The objective of this project is to develop efficient algorithms for finding discriminant functions, to investigate the classification capabilities of different pattern classifiers and to develop good procedures to extract features.

Progress Made During the Past Year:

(1) A very efficient method, called the accelerated relaxation method, for finding solutions of linear inequalities, has been invented, implemented and tested. Experimental results showed that this method is very efficient.

(2) The problem of determining the pattern classifying capability of machines using threshold logic is transformed into a formal logic problem. Some interesting results have already been found.

(3) Although there are many feature extraction procedures, none of them are very suitable for pattern recognition purposes. Recently, we have studied extensively the possibility of using Fisher's criterion. Two algorithms have been proposed, implemented and tested. This method has also been compared with the Kerhunen-Loeve expansion.

Proposed Course:

We shall continue working on the development of new pattern recognition techniques. We shall also investigate the possibility of using MULTIPLE to classify patterns.

Publications:

1. Chang, C.L.: The Accelerated Relaxation Method for Linear Inequalities. IEEE Trans. on Comp., Vol. C-20, No. 2, Feb. 1971, 222-225.

2. Hodes, L.: A Programming System for the On-line Analysis of Biological Images. Comm of the ACM, May 1970, 279-284.

3. Hodes, L.: Discrete Approximation of Continuous Convex Blobs. SIAM J. of Appl. Math., September 1970, 477-485.

4. Slagle, J.R. and Lee, R.C.T.: Application of Game Tree Searching Techniques to Sequential Pattern Recognition. Comm. of the ACM, February 1971, 103-110.

Serial No. DCRT 6.4

1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Chemical Notation and On-Line Information Systems

Previous Serial Number: Same

Principal Investigators: Deena A. Koniver, George Miller

Other Investigators: Richard Feldmann (CCB)

Cooperating Units: CCB

Man Years

Total: 2.4
Professional: 2
Other: 0.4

Project Description:

The immediate objective of this project is the automatic encoding and decoding of organic chemical compounds in Wiswesser Line Notation (WLN), a widely used linear notation. The overall objective of this project is to enable a chemist to easily communicate with computer-based chemical information systems.

Significance to Biomedicine:

This project complements related chemical projects in DMB and CCB to the extent that the WLN is an intelligible code and not just a random string of digits, as is the Chemical Abstracts Registry Number or a computer-based connection table. Much useful searching of a WLN file can be preformed without the aid of a computer.

Methods Employed:

There are two basic WLN computer programs being developed for the PDP-10 computer. One accepts as input the atom-by-atom connection table of a compound and generates the corresponding WLN. The input connection table can be obtained by several possible means, including a chemist (or user) drawing the structural diagram on the Rand Tablet/display. A computer program was written which transforms a Chemical Abstracts compact connection table into the redundant form of connection table used by the WLN generation program, thereby enabling us to process large files of compounds stored in Chemical Abstracts format.

The other basic program converts an input string of WLN to its corresponding connection table. Since a connection table itself is rather unintelligible, a program is then used to generate a display of the structural diagram from the connection table.

Current Status:

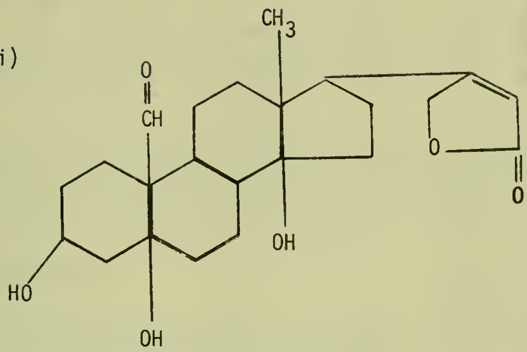
Both the encoding and decoding programs are able to handle, in general, compounds as complex as polycyclic fused ring systems, and even a chain of two such ring systems.

The WLN generation program can correctly encode about 95% of a set of published compounds tested for carcinogenesis, and about 81% of the compounds in the Common Data Base, a set of 20,000 compounds, jointly compiled by the Food and Drug Administration and the National Library of Medicine.

The following examples are all compounds which can be handled by both programs; i.e., given the structural diagram or connection table the WLN can be generated, and given the WLN the connection table and then the structural diagram can be generated.

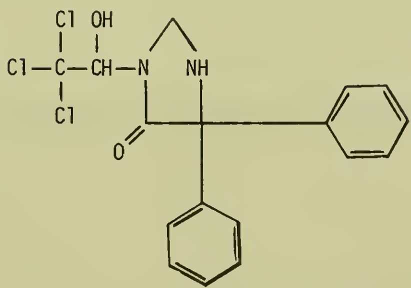
Examples:

(i)



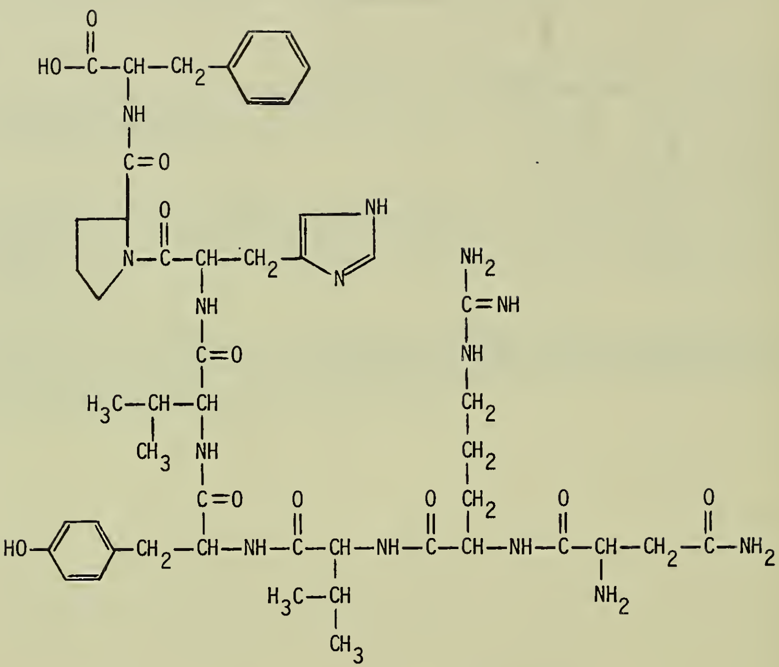
WLN: L E5 B666TJ AVH E IQ MQ OQ F- DT5OV EHJ

(ii)



WLN: T5NV DMTJ AYQXGGG CR& CR

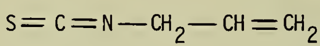
(iii)



WLN:

T5NTJ BVMYVQ1R& AVYMVYY&&MVY1R DQ&MVYY&&MVY3MYZUM&MVYZ1VZ1- DT5M CNJ

(iv)



WLN: SCN2U1

Proposed Course:

Currently neither program can handle a connection table with disconnected fragments, such as is found with salts, ions, etc. It is planned that both programs will be extended to handle such cases.

The feasibility of extending either or both of the programs to handle more complex situations, such as bridges and spiro rings, will be studied. Unless an actual user of the programs has a real need for the handling of such complexities, it is unlikely that the programs will be expanded much beyond their present capabilities.

Paper Published:

1. Farrell, C.D., Chauvenet, A.R., and Koniver, D.A.: Computer Generation of Wiswesser Line Notation. J. of Chemical Documentation, Vol. 11, No. 1, 1971, 52-59.

Serial No. DCRT 6.5

1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: On-line Interactive Graphic Curve Fitting System

Previous Serial Number: None

Principal Investigators: Gary D. Knott

Other Investigators: D. Reece, R. Schrager (PSL)

Cooperating Units: PSL

Man Years:

Total:	0.68
Professional:	0.58
Other:	0.1

Project Description:

The use of curve fitting to reduce laboratory data is a ubiquitous process. The objective of this project is to convert the graphic curve fitting system called MODELAIDE to the PDP-10. This will result in saving the rental of an IBM 2250 graphic display and will consolidate our graphic efforts at DCRT on the PDP-10. We are taking advantage of this conversion to extend considerably the capabilities of the system.

Progress Made During the Past Year:

Preliminary design work has been completed and major pieces of programming have been done.

Proposed Course:

Further programming is required, to be followed by appropriate debugging.

Serial No. DCRT 6.6

1. Heuristics Laboratory
2.
3. Bethesda

PHS - NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Computer System Support

Previous Serial Number: None

Principal Investigator: Robert Sproull

Other Investigators: Gary D. Knott, Douglas Reece

Cooperating Units: CSL, CCB

Man Years

Total:	1.08
Professional:	0.88
Other:	0.2

Project Description:

The objective of this project is to upgrade the currently existing service of the PDP-10 system.

Progress Made During the Past Year:

(1) With the cooperation of members from CSL and CCB, a high speed parallel (1 Megabit per second) link between the PDP-10 and the IBM-360 has been designed. This link was stimulated by the needs of the chemistry project. Subsequently, the PDP-10 can use 360's file system.

(2) Many software packages developed in Stanford University have been installed on our PDP-10 system. These include SOS, SAIL, and FAIL.

(3) An effort has been made to attract more scientists to use the PDP-10 display system. A program has been written to aid in measuring the lengths of replicated DNA strands. The user uses the RAND Tablet to trace the strand and the program provides immediate indications of the lengths of the strands.

(4) An investigation has been made as to the possibility of using a "half-line" printer for the 360. With this device and some other facilities added to WYLBUR, one can use WYLBUR to print subscripted and superscripted expressions.

Publications:

1. Feldman, J. A. and Sproull, R. F.: System Support for the Stanford Hand-Eye System. Proceedings of the Second International Joint Conference on Artificial Intelligence, 1971.

July 1, 1970 through June 30, 1971

PHS-NIH

Division of Computer Research and Technology

Summary of Branch Activities — 1. DCRT-7

7. Data Management Branch — 3. J. Emmett Ward, Branch Chief

I. SUMMARY

The Data Management Branch is responsible for providing NIH with efficient, effective and easily accessible computer-based capabilities for data management and analysis. Primary emphasis during fiscal year 1971 continued to be standardization; this included establishing new standard programming techniques and expansion of old approaches.

During this period, the DMB has added a new tool, the Recursive Macro Actuated Generator (RMAG), to provide the capability of generating source code in any language. Using RMAG, a programmer can generate many tedious repetitive coding requirements or entire source language programs.

With the capabilities of macro generation, the Branch has redesigned the standard update to provide users with complete update and edit capabilities in one COBOL program. The generated program provides the user with field name update, eliminates the need for extensive job control language and internal program calls and provides a machine-free generated source program. It is planned to extend this approach to generating file inquiry and analysis techniques.

Currently, the DMB is in the process of re-evaluating the Mathematical and Statistical programs available to the NIH. It is intended that a completely revised Math/Stat library will be made available to the scientific community by early FY72. A new set of documentation will be provided and this documentation will be revised and redistributed as additions, deletions and changes to the library occur.

As can be seen in the detailed reports following this summary, the Branch has provided many Institutes and Divisions with new computer based capabilities during this fiscal year. We feel that this Branch will become even more effective in the coming year as the systems capability of the Computer Center Branch expands and we are able to offer more efficient techniques for solving data management and analysis problems.

During recent months the Data Management Branch has implemented a procedure for converting the existing Clinical Center Pathology File data. This conversion opens the door to investigators in the Clinical Center. They may now search laboratory data which is endemic to their patients for the purposes of unique reporting and analysis. It is intended that during the coming fiscal year the Data Management Branch complete the necessary conversion of

old and new laboratory data, provide for recurring file maintenance and establish individual systems for each Clinical Center investigator.

It is also anticipated that the Data Management Branch will provide additional systems analysis and programming support for such areas as drug administration and pharmacy inventory control within the Clinical Center.

In providing support for chemical information processing at NIH, the Data Management Branch is making available such capabilities as substructure search and current awareness search of chemical literature. It is intended for the next fiscal year to extend current literature search capabilities and to investigate various means of providing interactive and batch processing capabilities for searching chemical structure information. This will include the following:

1. Establish continuing communication with the expected user community to define requirements, expected usage of the system and types of access.

2. Provide a means for efficiently searching the rather massive chemical structure files.

3. Implement for recurring production both current awareness and retrospective searching of the Chemical Biological Activity tapes.

4. Investigate new sources of chemical literature data for inclusion in the regular search package.

A. Organizational Developments.

During this year the internal organization of the Branch was revamped for two reasons:

1. To better define assignments of data processing tasks within the Branch.

2. To provide a better method of controlling documentation standards. This control will assure that each programmed system can, in fact, be operated according to its set of documentation guidelines.

The Mathematical/Statistical Unit was transferred into the Branch from the Laboratory of Applied Studies.

Serial No. 7.1
1. Data Management Branch
2. Applied Systems
   Programming Section.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Cigarette Condensate Study

Project Leader: G. Dobenecker

Other Professionals: None

Cooperating Units: Etiology, NCI

Man Years:

Total:	0.15
Professional:	0.15
Others:	0.00

Project Description:

Objectives:

The purpose of the system was to collect and report on information gathered by Hazelton Laboratories. The information pertained to the occurrences of tumors (size, number, kind, etc.) in 7,000 animals treated with applications of a nicotine type solution in varying concentrations. The specifics of the solution were determined by Dr. Gori of the National Cancer Institute.

Methods Employed:

Forms were designed for use by Hazelton Laboratories in the recording of the information. The information is submitted monthly to the Key Punch Section of DCRT. The punched information is appended to the specific animal records and eighteen one-page reports are generated. These reports are designed to show trends in the occurrence of tumors.

Major Findings:

Experiment still in progress.

Significance:

None as far as interesting or unique computer techniques employed.

Proposed Course:

The experiment is currently scheduled to run for eighteen months. At the end of the experiment a determination will be made as to the need or use for more detailed statistical reports on the accumulated data. Additional programming support will be supplied upon request by the Data Management Branch of the Division of Computer Research and Technology.

Honors and Awards: None

Publications: None

Serial No. 7.2
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Carcinogenesis Bioassay Data System Implementation

Previous Serial Number: 7.2

Project Leader: Mary E. Linhart, DCRT

Other Investigators: Larry Martin, DCRT
Sandra Foote, DCRT
Anne Gallagher, DCRT
John Parks, DCRT

Cooperating Units: Program and Data Analysis Unit, Office of the Associate Director for Carcinogenesis, Etiology Area, NCI

Man Years:

Total:	2.85
Professional:	2.80
Others:	.05

Project Description:

Objectives:

Implement the Carcinogenesis Bioassay Data System which will facilitate contract administration and permit more satisfactory evaluation of experiments. Design forms and input procedures for gathering standard compendia of information which are fundamental to carcinogenic studies. Provide computerized editing and file maintenance procedures. Develop a basic format to structure reporting of test results in all investigations. Develop procedures for user handling of input and systems operation.

Methods Employed:

Standard information system design and programming techniques have been used.

Major Findings:

Input forms and file records have been designed. Specifications for eleven major programs have been developed and programming is underway. Redesign of input forms and some related file records has forced revision of some programs. Initial program and system testing has been completed.

Significance to Bio-medical Research and the Program of the Institute:

The Carcinogenesis Bioassay Data System is intended to facilitate analysis of experimental results on large numbers of animals and tests by imposing structure upon massive collections of data so that comparisons and correlations can be made upon experimental results within an investigation and among multiple investigations. It is expected that establishment of files will allow for more sophisticated analysis of experimental results.

Proposed Course:

National Cancer Institute personnel will assume responsibility for data collection, file maintenance, and reporting of basic information.

DCRT will continue to provide maintenance and support of the basic systems during the coming year. In addition, other programs will be developed as desired by the Program and Data Analysis Unit to correlate data and report findings.

Honors and Awards: None

Publications: None

Serial No. 7.3
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: NIAID Research Contracts

Previous Serial Number: None

Project Leader: Judy Mahaffey, DCRT

Other Professionals: David Blessley, DCRT

Cooperating Units: NIAID

Man Years:

Total:	0.2
Professional:	0.15
Others:	0.05

Project Description:

Objectives:

The purpose of this project was to design and implement a system to 1) organize research contract data into a computer accessible file 2) provide file update capabilities 3) prepare the necessary recurring management reports and 4) provide the capability of retrieving, in an organized fashion, data needed to meet one-time management requirements.

Methods Employed:

A WYLBUR system with IRS capabilities has been designed and implemented. The system makes use of both Batch and on-line processing. All research contracts data is entered directly into the system through a 2741 terminal and stored on a WYLBUR data set. A COBOL program is then run periodically to edit and re-format this data for later processing. Errors identified by the COBOL program are corrected on-line. After the update cycle is complete, IRS and COBOL programs are submitted to prepare a variety of management reports.

Major Findings and Significance:

The system now serves as a powerful data management tool for NIAID. It stands as proof that a single non-data processing person can maintain and operate a computer system if the system is so tailored to the user. Perhaps the primary significance of the system is that all research contracts data can be maintained and selectively invoked for any one or all 24 report options by submitting a single computer run. In twenty minutes to two hours the requested reports are available to management. All time previously spent on manual data verification and report preparation is freed for analysis tasks.

Proposed Course:

The complete system and documentation package have been turned over to the Contract Management Branch, NIAID.

Honors and Awards: None

Publications: None

Serial No. 7.4
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: KURU Disease Computer System

Previous Serial Number: None

Project Leader: George Dobenecker, DCRT

Other Professionals: None

Cooperating Units: NINDS Collaborative and Field Research Branch

Man Years:

Total:	0.20
Professional:	0.20
Others:	0.00

Project Description:

Objectives:

The purpose of the project was to develop a computer system to record, report and maintain data, that had been and was being accumulated, relating to the KURU disease.

Methods Employed:

A card system had been used to accumulate the information so the initial step was to convert that data to a tape oriented file. New input documents were designed to make the new data compatible with the input requirements of the computer system. The system was supported by the standard update facility whereby records can be inserted, modified and/or deleted just by entering the proper updating code (I,M,D) and the key of the record in question. Report programs were written to produce listings of the file by name and by village.

Major Findings:

None

Significance:

The facility was created to easily and accurately accumulate, maintain, and report on information pertaining to the KURU disease.

Proposed Course:

The system as currently designed is maintained and used by NINDS field research. Additional programming support is supplied upon request by Data Management Branch.

Honors and Awards: None

Publications: None

Serial No. 7.5
1. Data Management Branch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Population Estimation for Inter-Census Years

Previous Serial Number: None

Project Leader: Anne Gallagher, DCRT

Other Professionals: None

Cooperating Units: Occupational Studies Unit, NIEHS

Man Years:

Total:	0.05
Professional:	0.05
Others:	0.00

Project Description:

Objectives:

In order to properly assess the effects of a suspected deleterious agent on an exposed population. In occupational studies, the comparison group commonly chosen is the general population of the state in which the "industry" is located. Since population figures specific for age-sex-race by state are not readily available, it is necessary to estimate inter-census populations.

Methods Employed:

Census data for 1960 is used as a basis for all estimations. Birth rates and the number of births for years other than 1960 are used to estimate the population for ages 15-44, since nearly all births are reflected in this age span. Likewise, for ages greater than or equal to 45, death data for years other than 1960 are utilized to estimate the population. Finally, the Current Population Reports Series P-25 are used. (P-25 series are population estimates by state and broad age groups released by the Bureau of

of the Census). Adjustments are made to the birth and death data population estimates by adding to them the amount they differ from the relative P-25 series values.

Major Findings:

Since the estimations have not been completed or utilized, there are no findings of major consequence as yet.

Significance:

The ready availability of this data will tend to increase the validity of studies using population comparisons.

Proposed Course:

Eventually, the population estimates will probably be updated with the information from the 1970 Census.

Honors and Awards: None

Publications: None

Serial No. 7.6
1. Data Management Branch
2. Scientific Applications Section.
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Longitudinal Studies Program of Human Physiology, Biochemistry, and Psychology (Aging Study): File Development Phase

Previous Serial Number: DCRT 7.6

Project Leader: Catherine W. Staneck, DCRT

Other Professionals: Personnel of Scientific Applications Section, DMB

Cooperating Units: Gerontology Research Center, NICHD

Man Years:

Total: 0.68
Professional: 0.58
Others: 0.1

Project Description:

Objectives:

The Gerontology Research Center of the NICHD has been collecting data related to the effects of aging from a continually changing population of about 700 active participants since 1958. In an effort to facilitate the analysis of this data it was decided to incorporate approximately 30 different data files associated with this project into a single masterfile having a single retrieval mechanism for all information in the project.

Methods Employed:

Initially a control file containing one record per person visit was established containing certain classification information for each individual such as date of visit and age. No data was admitted into the data masterfile unless a corresponding control record was present in the visit file. Each data record also received a preliminary edit

of the data fields prior to being included in the masterfile. Several general purpose file maintenance and retrieval systems available in the DCRT program library have been incoprorated into the data management process. A 2741 terminal has been installed at the Gerontology Research Center in Baltimore to enable the group there to control the updating of the masterfile, currently containing twenty different data types, directly by submitted batch jobs through Wylbur. Data retrieval is being accomplished through the use of a program, written specifically for the project, which selects specified data from the masterfile containing more than 30 different data formats and produces a standard formatted output for use in statistical analysis routines.

Major Findings:

Basic statistical analysis of diverse data can be a useful editing tool in identifying errors which have resulted from the data collection process. Results lying outside two standard deviations from the mean for a particular variable can be checked where a detailed manual edit of the data on a field-by-field basis is impractical. The data types in this project contain as few as three fields or as many as 106 data fields. To identify problems within the set of records for an individual a regression analysis was also incorporated into the project.

Significance:

Usually data editing and validation are a preliminary step which must precede any statistical analysis of the data. This project has shown the usefulness of incorporating basic statistical measurements into the edit process itself. This is especially true when a large volume of varied data which has been collected over a long period of time is being processed.

Proposed Course:

The Gerontology Branch is now updating and retrieving data from 20 data types currently in the masterfile. The means and regression computations along with other report features have been programmed in such a way as to make data selection and analysis possible for someone with minimal computer experience. The programs and procedures are currently being executed by use of remote processing from Baltimore.

Three of the data files which require special editing are still being handled by DMB. And four of data files which

are not compatible with the identification process used in matching data records with the visit control file are inactive.

Honors and Awards: None

Publieations: None

Serial No. 7.7
1. Data Management Branch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Life Tables and Chi-Square for Epidemiological Studies

Previous Serial Number: None

Project Leader: Merle Powell, DCRT

Other Professionals: None

Cooperating Units: Occupational Studies Unit, NIEHS

Man Years:

Total:	0.53
Professional:	0.53
Others:	0.00

Project Description:

Objectives:

The Occupational Studies Unit conducts studies of the relationship between chronic disease mortality and prior occupational history. DMB has supported this effort by creating and maintaining data files and supplying computer programs as requested. The programs described in this report were written to provide the statistical support for a series of papers examining work-histories and mortality in specified areas in the steel industry.

Methods Employed:

"Expected deaths in each work area for specified race, age, nativity and residence were calculated by applying the specific rate of the total steelworker population to the number at risk in the work area. Expectations over various levels of these characteristics were then obtained by summation."

Significant differences were determined by a one-degree of freedom summary chi-square. The method used in the program is from "Extensions of the Mantel-Haenszel Procedure", Nathan Mantel, Journal American Statistical Association, Vol. 58, pp. 690-700, September 1963.

Major Findings:

See the Journal of Occupational Medicine for the series of articles.

Significance:

The life table analysis programs and the chi-square program were written in a general manner so that they would be applicable within a wide range of similar studies.

Proposed Course:

The study will be continued as time and money resources allow.

Honors and Awards:

On April 21, 1971, Dr. Lloyd will receive the Adolph G. Kammer Merit in Authorship Award for the two papers of the series supported by these programs which were published within the time period considered by the Award Committee.

Publications:

A series of papers is being published by Dr. Lloyd et al in the Journal of Occupational Medicine.

Serial No. 7.8
1. Data Management Branch
2. Software Support Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Generated Update Facility

Previous Serial Number: None

Project Leader: Robert A. Magnuson, DCRT

Other Professionals: Joseph Campbell
Barry Madia
Darius George

Cooperating Units: None

Man Years:

Total:	02.3
Professional:	02.3
Others:	00.0

Project Description:

Objectives:

Provide source-program generators such that the generated programs are the tools for establishing and maintaining a computer file with virtually no programming effort. Minimize the time and effort for converting users from manual and punch card systems to computer tape and disk systems. Align the use of this function with the existing Standard Update Facility and with the Query system, thus, providing the user with the means of querying and reporting on the data he has created.

Methods Employed:

Source-program-generation techniques were applied to the existing DMB Standard Update Facility, a (fixed) catalogued procedure which required the special writing of a new, ad hoc edit procedure for each different intended file. Control for each input transaction is specified by a fixed-format change record containing an

action code (insert, modify, or delete), key, numerical position and length target field specifiers, and the new information.

The conversion from the Standard Update Facility to the Generated Update Facility was attacked in two steps, EDITGEN and UPDATEGEN. EDITGEN, a program which generated a COBOL source edit program for use with Standard Update was first implemented using the RMAG language. Requiring that the intended file's fields be simply described, and that that set of file descriptions be used as input, a complete COBOL edit program was automatically generated. Three other versions of EDITGEN were then implemented in parallel--one in COBOL source, one in PL/1 source, and one in ALC source. Each of these four versions, using (logically) the same input data (i.e., the field descriptions), generates the same COBOL source edit program. But the coding techniques used, the lapsed time to implement, and the cost to generate the COBOL program for a given file structure, varied along predictable lines.

UPDATEGEN, a program which generates a complete COBOL source update program, was first implemented in ALC--using portions of the ALC EDITGEN. The user now has much more freedom and gets a better, cheaper running Update System. The input for UPDATEGEN is once again the file description. Whereas the Standard Update required the key in a fixed position, UPDATEGEN accepts it wherever desired. Data input preparation for the Update Run is made much easier since the input can be in free form. Only the fields that are changed are checked. The fields are given names thus eliminating the need to indicate position and length.

The trade-offs were interesting. An RMAG implementation could be written and debugged in several days, whereas the other language implementations took several weeks. The cost to generate went the other way--less than a dollar for the non-RMAG generators to generate a large edit, and $6.00 for the RMAG generator to generate the same COBOL program. It would appear that since each COBOL program is generated but once, the language to use for the generators is RMAG.

Significance:

The COBOL programs generated by these systems were large, comprehensive systems. They did not appear to have been automatically machine-generated but looked, rather, like

they had been written by human programmers. In the event that (say) an edit requirement was such that the generator had not been designed for that option, it was relatively easy to modify the generated COBOL source program to accommodate it.

Proposed Course:

The system may be extended for a larger set of requirements allowing for more complicated editing logic.

Honors and Awards: None

Publications: None

Serial No. 7.9
1. Data Management Branch
2. Software Support Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Recursive Macro Actuated Generator

Previous Serial Number: None

Project Leader: Robert A. Magnuson, DCRT

Other Professionals: None

Cooperating Units: None

Man Years:

Total:	0.75
Professional:	0.75
Others:	0.00

Project Description:

Objectives:

Provide a very powerful programming tool for the generation of source programs, text, generators, and arbitrary strings.

Methods Employed:

Based on the principal investigator's previous work, a stand-alone macrogenerator was implemented in OS/360 ALGOL. It was called RMAG and could be run via WYLBUR or via production batch. The appropriate JCL procs were written and catalogued. A library of RMAG subprograms was implemented. These programs were used as part of various RMAG applications later programmed. Comprehensive documentation was prepared including a programmed primer, a glossary of terms, an annotated explanation of the library, and a summary. In order to decrease the cost of running RMAG programs, the entire system was reimplemented in OS/360 ALC. Significant new features based on previous use were also included. The library was recoded based on the new features. The documentation was completely rewritten and reissued.

Major Findings:

The RMAG language was successfully used in a number of projects, e.g., Generated Update Facility, by the principal investigator. In addition, other DMB, DCRT, and Institute scientists used RMAG on their projects. The ALC version reduced the cost of running RMAG programs to a small fraction of that of the ALGOL version. (A typical run cost one-seventeenth or less.)

Significance:

An RMAG Source program is independent of running computer, target computer, and target language. E.g., SEL 810 ALC has been generated using RMAG on the 360. RMAG is being used to generate ALC, COBOL, FORTRAN, ALGOL, JCL itself, and other public as well as private-language programs.

Proposed Course:

The system will be extended and even more broadly applied. Of particular power is the ability to generate generators, to generate such generators, etc. Useful applications will be sought and implemented.

Honors and Awards: None

Publications: None

Serial No. 7.10
1. Data Management Branch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Current Awareness Search of Chemical Literature

Previous Serial Number: None

Project Leaders: S. C. Foote
S. R. Heller
K. P. Shapiro

Other Professionals: None

Cooperating Units: NIAMD (Dr. Keatha Krueger)

Man Years:

Total: 1.0
Professional: 1.0
Others: 0.0

Project Description:

Objectives:

The development of a bi-weekly current awareness searching of the CBAC (Chemical Biological Activity) Journal published by Chemical Abstracts Service.

Methods Employed:

Text-search Programs obtained from the National Science Library of Canada were modified to search CBAC tapes in MARC format. Programs to reformat the CBAC tapes into the proper format and to print results were written. MARC format is the Library of Congress Standard.
Results: Testing of approximately 200 profiles submitted by volunteer users is in progress.

Significance:

Provides chemists at NIH with a searchable data base of selected literature on chemical compounds of biological interest.

Proposed Course:

A fully debugged system to be run by recurring production has been implemented and tested. Following development of a charging algorithm, DCRT should realize a profit from the project.

Honors and Awards: None

Publications: None

Serial No. 7.11
1. Data Management Baanch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June,30, 1971.

Project Title: Evaluation of Dermatoglyphics Data

Previous Serial Number: None

Project Leader: Louise Saulnier, DCRT

Other Professionals: None

Cooperating Units: Children's Diagnostic & Study Branch, NICHD

Man Years:

Total:	0.33
Professional:	0.33
Others:	0.00

Project Description:

Objectives:

Aid in the evaluation of finger and palm print data of several distinct populations. The analysis consists of obtaining cross-frequency tables comparing left hand versus right hand statistics, obtaining counts and percentages of certain data fields, and performing standard statistical calculations on those data fields with continuous distributions.

Methods Employed:

Thus far, two large populations have been analyzed. Certain edit routines have been developed which are common to whatever type of data is being processed, but the type of populations and the analysis required for each have made it necessary for specialized programs to be developed for each analysis. The first analysis was a study of a control population from Lancaster County,

Pennsylvania, to be used in a Cleft Palate Study. Comparisons of left and right hand data were made, as well as analysis of data within a patient's family. The second study was used to develop generalizations about the dermatoglyphic features of Micronesea populations. Counts, percentages and statistical tabulations were performed.

Major Findings:

Computer analysis has aided the geneticist requiring the programs from having to perform calculations which were done previously with a desk calculator.

Proposed Course:

The geneticist requiring the programming is still feeling his way as to what he can obtain with computer analysis. His goals have not been definite enough to develop any specifications for a more consolidated system, but he expects to be submitting more data as it becomes available.

Honors and Awards: None

Publications: None

Serial No. 7.12
1. Data Management Branch
2. Mathematical/Statistical Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: The Effect of Secobarbital on Family Interaction

Previous Serial Number: None

Project Leader: Dave Van Sant

Other Investigators: None

Cooperating Units: Section on Experimental Group and Family Studies, NIMH

Man Years:

Total: 0.25
Professional: 0.25
Others: 0.00

Project Description:

Background:

This was a study designed by NIMH to determine the stability of patterns and styles of family problem solving. The challenge to this stability was 175 milligrams of secobarbital given to the offspring of half the families in the study. These were compared with a group of families where the child received a placebo. Secobarbital was administered to produce an acute cognitive and perceptual deficit in the child.

Objectives:

Programs were written to evaluate the results obtained from both groups and to show if there was a statistically significant deterioration in the drugged group.

Methods Employed:

Evaluation of data in this project consisted of analyses of speech patterns as they were coded from a polygraphic record of vocalizations by all family members in the experiment (computing speech rates, average statement length rates, interruption rates ... etc.) and analyses of variance and correlations that permitted the determination of the more precise nature of the drug effect.

Major Findings:

As evaluated by NIMH, Significance for mental health: The study suggests that a psychoactive drug may have very limited effect on the family constellation; this has implications for the use of psychoactive drugs in family-oriented therapy programs. It also provides more basic data on the family as a problem solving group and gives some index as to what may or may not determine how the family functions in that capacity.

Honors and Awards: None

Publications: None

Serial No. 7.13
1. Data Management Branch
2. Mathematical/Statistical Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Generalized Clinical Chemistry Data Handling System

Previous Serial Number: None

Project Leader: George Shakarji

Other Investigators: Kirk Dorn

Cooperating Units: None

Man Years:

Total:	0.6
Professional:	0.6
Others:	0.0

Project Description:

Objectives:

A set of programs were developed for editing, storing, updating, retrieving and analysis of clinical laboratory data. The programs handle variable as well as fixed length records of demographic information and test results linked to form patient files. Optional statistical packages include examination of time trends and analyses of variance to isolate intra-individual or inter-individual components of variance. Applications of these programs produces a patient file consisting of a demographic record followed by a chain of test records, each containing the results of a laboratory analysis of a given sample. Further responsibilities of the system concern the maintenance of the patient file by a) adding new demographic and test records to the patient file: b) making appropriate changes and deletions: c) deleting spurious

information which may even be syntactically correct and therefore pass the edit program.

Methods Employed:

Primary keys link demographic and test records forming the patient's file. Secondary keys link the separate laboratory tests performed on the same sample of body fluid. Each demographic record contains an updatable counter indicating the number of test records (i.e., sample accession numbers) stored for that particular patient: minimum of one test record with a high upper limit. Each test record contains a similar updatable counter, telling the number of test results following in that record. Provision is made for changes in analytic method at any point in the time series of test results.

Significance:

Application of the edit and update program produces a patient file that contains the results of laboratory analysis of given samples. The patient file provides a data base for scientific evaluations in biomedical research areas. Procedures for building each patient file were designed in a form suitable for statistical analysis. The statistical programs, included as options in the present program package, analyze each patient file (or control series) as a separate entity, they should be immediately useful to the reduction of a large proportion, if not all, of the data in most clinical studies.

Proposed Course:

These programs are currently being applied to produce files for patient data summary records for the Endocrinology Branch of the NHLI. Further uses of these programs for clinical chemistry data look promising.

Honors and Awards: None

Publications: None

Serial No. 7.14
1. Data Management Branch
2. Mathematical/Statistical Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Sequence Analysis of Mother-Infant Interaction

Previous Serial Number: None

Project Leader: Ray Danner

Other Professionals: None

Cooperating Units: Child Research Branch, NIMH

Man Years:

Total:	0.2
Professional:	0.2
Others:	0.0

Project Description:

Background:

As a part of a longitudinal study of family development and parent-child interaction conducted by Child Research Branch, NIMH, continuous three-hour observations of mother-and-offspring behavior were made at specific intervals between the first and fourteenth weeks post partum. The observations indicate which responses occurred at any point in time, how long they lasted, and the behavioral contexts in which they occurred, as well as which behaviors preceded or followed them.

Objectives:

It is hoped that the statistical programming done at DCRT will enable the Child Research Branch to evaluate the statistical significance and behavioral magnitude patterns of inter-and-intra-individual behavior.

Work done in FY 1971:

The development and application of programs involved the following steps:

1) Ordering the sequence of elements in each of approximately 73 individual decks and editing the data for internal consistency within each deck.
2) For every response-class within each deck, the determination of a frequency distribution of bout-durations (the duration of each separate occurrence of the response).
3) The development of a procedure whereby different bout-durations within a response-class may be treated as different response classes.
4) A Monte-Carlo shuffle of response-bouts in terms of frequency and duration to indicate the nature of "random" relationships between behavioral events.
5) An analysis of the behavioral data contained in each deck yielding a frequency-distribution of the time-intervals between:
   (a) Onset of pairs of response-classes for all possible response pairs.
   (b) Onset of one and termination of another response-class for all possible response pairs.

Major Findings:

This study is not yet complete. Thus major findings are not available at this time.

Proposed Course:

Assuming the foregoing evaluations are adequate, a chi-square analysis is proposed by the Child Research Branch for the evaluation of inter-cell frequencies. Should this test reveal non-chance patterns of response, it would be desirable to have a measure of the amount of variance accounted for the predictability afforded by a knowledge of pattern.

Honors and Awards: None

Publications: None

Serial No. 7.15
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: NIH Space

Previous Serial Number: None

Project Leader: Felix Liski, DCRT

Cooperating Units: Space Management Section, ODA, A, PO

Man Years:

Total:	0.32
Professional:	0.00
Others:	0.32

Project Description:

Objectives:

This system was developed to establish a computer-based update system to replace the manual punched card system for the Space Management Section. Initially, the punched card master space file was edited and converted to a magnetic tape file, modified to conform with the prescribed WYLBUR format. This system will capture and prevalidate all input data transactions for updating the space master magnetic tape file. It will also prepare reports to account for all space for each room and floor in each building on the reservation and in rental buildings, as well as space occupied by personnel at field stations.

Methods Employed:

All input data is captured at tele-communications terminal located in the Space Management Section office, and kept in on-line files until used by the update segment of the system. The system combines Batch processing, using COBOL, RPF, and WYLBUR.

Major Findings:

1. Data can be easily and efficiently entered by non-data processing personnel in Space Management.

2. Updating of master records is keyed on five position (serial number) for identification of each record as opposed to 23 position key consisting of organization, building, floor, room, and purpose codes. Therefore, few mistakes occur and less time is spent on the terminal. The system became operational July, 1970.

Significance:

Capture of all input data is done in Space Management on the terminal using the WYLBUR system. All computer programming is performed in DCRT.

Proposed Course:

A revised system should be devised using package programs to be turned over to Space Management Section to make them independent of DCRT.

Honors and Awards: None

Publications: None

Serial No. 7.16
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: NIH Committee Management

Project Leader: Penny Brogan, DCRT

Other Professionals: Jeanne Arlotta, DCRT
Paul Boone, DCRT
Clinton Stoops, DCRT

Cooperating Units: Committee Management Office, OD
Systems Application Section, DCRT

Man Years:

Professional:	0.58
Others:	0.00
Total:	0.58

Project Description:

Objective:

The purpose of this project is to design and implement a system for maintaining a computer file with the basic information on members of NIH Public Advisory Committees. The main byproduct of this system was a roster of the committee members including indexes and table of contents. Various master and selective listings were also outputs from the system.

Methods Employed:

Data for the system was collected on code sheets and later key-punched. This method was selected instead of WYLBUR because it was easier and required less time for the Committee Management Office. The cards were used to update the master tape file monthly. The eleven programs used by the system were written in COBOL.

Significance:

The system saves a lot of time in the record-keeping process for the Committee Management Office.

Additional Uses:

DHEW decided to set up a computerized committee management system at department level. The program that produced the NIH Roster was modified to produce a DHEW Roster.

Maintenance:

All programs and documentation were turned over to the Systems Application Section (Clinton Stoops) for maintenance.

Honors and Awards: None

Publications: None

Serial No.7.17
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Framingham Heart Study, Exam 10 Inventory

Project Leader: Felix Liski

Other Professionals: Evelyn Knight, DCRT

Cooperating Units: Biometrics Branch, NHLI

Man Years:

Total:	0.25
Professional:	0.25
Others:	0.00

Project Description:

Objectives:

This system is designed to produce descriptive frequency tables of symptoms and medical results of the tenth set of exhaustive medical examinations given to an established cohort of males and females living in the area of Framingham., Mass. Completed reports are evaluated by sponsor, Dewey Shurtless, Statistician, Biometrics, NHLI.

Methods Employed:

Input data was initially keypunched onto tape and edited by the sponsor. Reports are prepared using the Tablemaker Language. For certain reports source data are extracted from two separate tapes for correlation.

Major Findings:

N.A.

Proposed Course:

Processing of the examination data will begin as soon as it arrives.

Honors and Awards: None

Publications: None

Serial No.7.18
1. Data Management Branch
2. Applied Systems
   Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Leukocyte Infusion Update and Reporting System

Project Leader: G. Dobenecker, DCRT

Other Professionals: Roger Dailey, DCRT
David Blessley, DCRT

Cooperating Units: Systems Analysis and Planning Office
Chemotherapy, NCI
Leukemia Service, MB, NCI

Man Years:

Total:	0.50
Professional:	0.35
Others:	0.15

Project Description:

Objectives:

The purpose of this system is to collect and report on Leukocyte Infusion information. The information collected included specific infusion information, laboratory follow-up information, and infusion series summary information. Prior to reporting on the information collected by our system we had to develop the facility to extract information from two related systems (i.e., the Cell Collection System and the Donor Recipient Information System).

Methods Employed:

Since the forms had been designed, prior to our entry into the project, we wrote a conversion program to make the source information coded and punched from the original forms compatible with the standard update input requirements. This was done to give the user the ability to insert, modify, and/or delete records using the standard updating approach in addition to the ability to enter data by coding the original source

forms. Prior to reporting on the information, a program is executed that combines the related records of the other systems and creates a combined file. The reporting programs address the combined file, thereby having all the information collected and maintained by the other systems available for the specific report requirements. A detailed infusion report was written. A summary report of infusion series information was written in addition to a laboratory follow-up report.

Major Findings:

Experiment still in progress.

Significance:

None as far as interesting or unique computer techniques employed.

Proposed Course:

When the system fulfills the collection and reporting requirements of Dr. Graw it will be maintained and executed by personnel of NCI. Additional programming support will be supplied by the Data Management Branch of the Division of Computer Research and Technology.

Honors and Awards: None

Publications: None

Serial No. 7.19
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title:	Great Lakes 1969 Influènza Study
Previous Serial Number:	None
Project Leader:	Felix Liski, DCRT
Other Professionals:	Margaret D. Myers, DCRT
Cooperating Units:	Division of Biologics Standards

Man Years:

Total:	0.17
Professional:	0.17
Others:	0.00

Project Description:

Objectives:

Purpose of the project is to compare the quality and results of the AICHI influenza vaccine manufactured by different firms, compare quality and results of two different methods of production. Comparison of antibody titers were made of the blood samples before and after the AICHI vaccine was given. 2834 volunteers participated in this study. Prior to receiving a vaccine, a blood sample was taken from each person to determine antibody titers in 16 antigen strains. Then each received the AICHI influenza vaccine. Subsequently, blood samples were taken again and measured. The comparison of the blood samples in the same individual showed a rise, a decrease, or no change in the antibody titer for a given antigen.

Preparation of descriptive computational tables were;

a) numbers and percent of volunteers exhibiting antibody responses by vaccine group, and by antigen.

b) geometric titers.

Methods Employed:

Input data was initially card punched and loaded onto magnetic tape, and edited for consistency, and for verification of manual arithmetic. For safekeeping, corrections were made in punched cards and then reloaded onto tape. COBOL program was used for editing and reformatting input records. Frequency tables were prepared using COBOL, Tablemaker languages, and WYLBUR. FORTRAN and WYLBUR were used to prepare geometric titer tables.

Major Findings:

N.A.

Proposed Course:

It is recommended that the sponsors continue to perform manual arithmetic on input data. The edit computer program will crossfoot where necessary and validate manual calculations.

Honors and Awards: None

Publications: None

Serial No. 7.20
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Parking Permit System (Assignment, Distribution and Maintenance)

Previous Serial Number: None

Project Leader: George Dobenecker, DCRT

Other Professionals: Roger Dailey, DCRT
David Blessley, DCRT

Cooperating Units: Protection and Parking Section, ODA, A, PS

Man Years:

Total:	0.25
Professional:	0.20
Others:	0.05

Project Description:

Objectives:

The purpose of the project was to assign permits to all the vehicles which may normally be parked on the reservation, and develop a system to maintain accurate information reflecting the current permit assignments.

Methods Employed:

Preprinted cards were punched for all employees on the personnel master file as of November 28, 1970. The cards were distributed by timekeeper number. The employees who received a card entered their building, room and phone number in addition to the license number of each car they might normally drive to work. Those employees who did not receive a pre-punched card, Form NIH-1234 (Rev. 9-70) provided the necessary information. The coded information was punched and appended to the parking permit master file by Social Security Number. Permits were assigned to specific license numbers and distributed to the employees.

The system is maintained by use of the standard update approach whereby records can be inserted, modified and/or deleted just by entering the proper update code (I,M,D) and the Social Security Number of the employee whose record is to be changed.

Report programs were written to produce listings of the file by employee name, license number, and permit number.

Major Findings:

None

Significance:

Parking Section was given the facilities to easily and accurately monitor those vehicles which may normally be parked on the NIH Reservation.

Proposed Course:

The system, as currently designed, will be maintained and used by Parking Section. Additional programming support will be supplied as requested.

Honors and Awards: None

Publications: None

Serial No. 7.21
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Type II Coronary Intervention Study

Previous Serial Number: None

Project Leader: Judy Mahaffey

Other Professionals: None

Cooperating Units: Biometrics Research Branch, NHLI

Man Years:

N.A.

Project Description:

Objectives:

The NHLI is in the midst of planning a study to determine if lowering the cholesterol level will cause a regression or decreased progression of premature coronary artery disease. The objective of this project is to design a complete and comprehensive computer-based system for the collection, storage and analysis of this data. The scope of this project will include the design of data collection forms, the development of data entry techniques, the design of file structures, and the development of data retrieval procedures. The system will encompass the functions of editing, updating, monitoring and reporting. The system is being designed so that it can be operated by personnel in the NHLI.

Methods Employed:

The project is currently in the preliminary systems design phase.

Major Findings:

N.A.

Significance:

N.A.

Proposed Course:

Complete forms design and initiate a total systems design phase.

Honors and Awards: None

Publications: None

Serial No. 7.22
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Railroad Retirement

Previous Serial Number: None

Project Leader: Judy Mahaffey, DCRT

Other Professionals: Penny Brogan, DCRT
Felix Liski, DCRT
Charles Twigg, DCRT
Vivian Fykes, DCRT
Jeanne Arlotta, DCRT

Cooperating Units: National Cancer Institute

Man Years:

Total:	1.2
Professional:	.9
Others:	.3

Project Description:

Objectives:

The objective of this project is to make it possible to do statistical processing on the Railroad Retirement data by creating a single, standard data file from the existing incongruent files which have evolved over the past ten years. This requires a complete analysis and redesign development of programs to convert the existing files and complete documentation.

Methods Employed:

The Railroad Retirement data files, consisting of over two million records in all, have evolved over the past years and have resulted in 7-9 different data formats and coding structures. In the systems analysis phase, all existing coding structures and formats were completely documented and correlated. A standard format, common coding scheme, and data validation method was developed for all files to be converted. A series of COBOL programs were executed to accomplish these tasks. Check digit procedures and control records were established to insure accuracy.

Significance:

The effort described above will make it possible for NCI to be a meaningful statistical analysis of the Railroad Retirement data. This design will facilitate analysis by tailoring processing to be compatible with the DCRT library programs.

Proposed Course:

When completed, a fully documented system will be turned over to NCI for independent statistical analysis.

Honors and Awards: None

Publications: None

Serial No. 7.23
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: URBS Systems Study

Previous Serial Number: None

Project Leader: Judy Mahaffey, DCRT

Other Professionals: None

Cooperating Units: Unit for Research into Behavioral Systems, NIMH

Man Years:

Total: .08
Professional: .08
Others: .00

Project Description:

Objectives:

The purpose of this project was to aid the Unit for Research into Behavioral Systems, CBRY, NIMH in establishing a computer system for statistical analysis of the Population Study data they have collected over the last few years.

Methods Employed:

A systems-approach to data analysis and file design was presented to URBS personnel. These techniques were implemented by URBS personnel in 1) defining their information goals, 2) evaluating their current data collection procedures and 3) defining and documenting their current status of their data. Guidance was provided through these stages. The project is not complete.

Significance:

Failure to establish a systems-approach in the initial stages of a data-collection project reduces the value of the data once it is collected.

Proposed Course:

A computer-based data storage and retrieval system will be proposed.

This will include file structures and data coding techniques. Methods for converting existing data to the computer data base will be outlined. Guidelines for documentation, control procedures and future data collection will be presented. Computer statistical processing techniques available through DCRT will be introduced.

Honors and Awards: None

Publications: None

Serial No. 7.24
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Histocompatibility Computer Analysis

Previous Serial Number: None

Project Leader: Claudette Thompson, DCRT

Other Professionals: None

Cooperating Units: Leukemia Service, M.B., NCI
Systems Analysis and Planning Office, Chemotherapy, NCI

Man Years:

Total: 0.40
Professional: 0.38
Others: 0.02

Project Description:

Objectives:

The objectives of this design are to allow for computer update and correction of data, to facilitate data retrieval to provide for analysis in evaluating test results, to mechanize the conversion of the raw data to the present data base, and to educate personnel in the use of the system.

Methods Employed:

Standard Information Analysis techniques have been used. Programs were written to eliminate the time-consuming manual workload, while analysis of the system was being done. Design, and most of the programming for a file maintenance system has been completed.

Significance to Bio-medical Research and the Program of the Institute:

This system is intended to aid in accurate HL-A typing, which is necessary to determine acceptable white cell and platelet transfusions and bone marrow transplants for leukemia patients. It is believed that the establishment of files will allow more sophisticated analysis of the experimental results.

Proposed Course:

Implementation of the system will be completed by the end of this fiscal year. After necessary debugging, the file maintenance system will be run in parallel with the present method of operation. During this period of time, personnel will be taught how to run the system. When the new file maintenance system is operating smoothly, programming for the analysis of the data will begin.

Honors and Awards: None

Publications: None

Serial No. 7.25
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: NIMH Publication

Previous Serial Number: None

Project Leader: Vivian Fykes, DCRT

Cooperating Units: Psychopharmacology Branch, NIMH

Man Years:

Total: 0.08
Professional: 0.00
Others: 0.08

Project Description:

Objectives:

The system is designed to create a data set on magnetic tape for input to the GPO Master Typography Program from a WYLBUR data set containing the required locators. The system allows data to be input in text form with indented, centered and/or multiply spaced lines.

Methods Employed:

An assembly language program concatenates the non-blank portion of lines and writes records with record format undefined and a block size of 3000.

Significance:

The system allows non-data processing personnel to input data in text form seeing some semblance of their final printed output. Data editing and validity checks are facilitated using the capabilities of WYLBUR.

Proposed Course:

The system became operational October, 1970 and has been documented and turned over to NIMH for future publications.

Honors and Awards: None

Publications: None

Serial No. 7.26
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Emergency Virus Isolation Facility Medical Monitoring Project

Project Leader: John Parks, DCRT

Other Professionals: None

Cooperating Units: Biohazards Containment and Control Section, VO, ET, NCI
Employee Health Service Branch, CC
Computer Center Branch, DCRT

Man Months:

Total:	2.0
Professional:	0.75
Others:	1.25

Project Description:

Objectives:

This system is designed to convert E.V.I.F. employee medical records to a computer-based file system and to provide basic maintenance and reporting capabilities for the file.

Methods Employed:

The DMB Standard Update System was used for file creation and maintenance. Input is key-punched from special pre-printed forms. Specific editing and reporting are handled by a single COBOL program using multiple entry points.

Major Findings:

Operational testing has revealed system maintenance problems which are relative to the nature of the data. Due to the personal nature of medical information, a private forms delivery procedure and manual separate of forms containing names and addresses is used to protect patient (employee) identity.

In summary, the system is not yet fully operational due to input volume, timing, and error-correction consideration and personnel

shortage.

Proposed Course:

For the immediate future, modifications to the system to allow transaction input and transaction error output to be WYLBUR data sets will facilitate system maintenance by reduced card handling and easier error correction. Additional changes to the system will make it user operable. These changes will provide for direct input to WYLBUR to facilitate data validation and correction. The WYLBUR data set is used by the system for input and the output of all rejected transactions for subsequent correction and collection. Also a step will be added to the system which modifies the data set containing the run JCL to rotate the tapes for the next run.

Honors and Awards: None

Publications: None

Serial No. 7.27
1. Data Management Branch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Psychophysiological Measures

Previous Serial Number: None

Project Leader: M.L. Dante, DCRT

Other Professionals: None

Cooperating Unit: LSES, IR, DCBR, NIMH

Man Years:

Total:	0.08	
Professional:	0.08	(Dante)
Others:	0.00	

Project Description:

Objectives:

To develop a generalized computer system for validation, storage, retrieval, and statistical analysis of psychological and physiological measures.

Methods Employed:

An analog tape which was produced during a psychological testing period is converted to digital. This digital tape is then converted to a standardized format and the measures time-locked and separated into individual disk or tape data sets. Each measure is edited, artifacts removed, and appropriate values computed. This information is then stored for subsequent retrieval and analysis.

Major Findings:

The system is not complete.

Significance:

As now being performed, the analysis of this type of data involves laborious and time-consuming hand-measuring and counting.

Proposed Course:

A flexible retrieval program is being designed which will select information from the files for user-specified time units, compute certain statistics, and output the data in a form suitable for further computer analysis.

Ultimately the system will be operated by the researcher with a minimum of programmer guidance.

Honors and Awards: None

Publications: None

Serial No. 7.28
1. Data Management Branch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title:	Uranium Miner Study
Previous Serial Number:	None
Project Leader:	Catherine Staneck, DCRT
Other Professionals:	Mary Lee Dante, DCRT Myrtle Morris, DCRT
Cooperating Units:	Occupational Studies, Unit, NIEHS

Man Years:

Total:	0.65
Professional:	0.65
Others:	0.00

Project Description:

Objectives:

For the past several years a continuing Federal study has been in progress to determine the effects of radiation exposure on Uranium Miners. To provide support for this investigation several sets of computer programs were developed for the H-800. These programs were converted for the IBM/360 and restructured, expanded, and generalized to fulfill additional requirements of the investigators.

Methods Employed:

In the past year a major updating of the data files used in the analysis has been accomplished. Several different types of mortality statistics were produced by modifying existing programs, giving them increased flexibility. In addition, several special computational programs were developed for the study.

A new system of programs was developed as a model for predicting the effects of exposure beyond the period of observation.

Major Findings:

Analysis of the information provided by this project has been handled by the Occupational Studies Unit, NIEHS.

Significance:

This project provided data management and analysis for the Federal effort to establish more precisely the hazards of uranium mining and to set safety standards.

Proposed Course:

Continued programming support of the project as it is required.

Honors and Awards: None

Publications:

A report was made by the Occupational Studies Unit to the Interagency Review group on Uranium Mining through the Special National Academy of Sciences Advisory Committee to the Federal Radiation Council.

The Surgeon General has requested that a PHS monograph be prepared based on this report.

Serial No. 7.29
1. Data Management Branch
2. Scientific Applications Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Cancer Survival System

Previous Serial Number: None

Project Leader: Catherine Staneck, DCRT

Other Professionals: Frances Bauer, DCRT
Lee McBride, DCRT

Cooperating Units: End Results Section, Biometry Branch, NCI

Man Years:

Total:	0.60
Professional:	0.60
Others:	0.00

Project Description:

Objectives:

The Survival System was originally developed six years ago to support the End Results Section, NCI, in preparation of the End Results in Cancer Reports. Maintenance and improvement of that system is the recent primary goal of the project, including both programming and documentation efforts.

Methods Employed:

The documentation has been extensively rewritten over the past year to enable a person not already familiar with the End Results method for calculating survival probabilities to make use of the system. In the original development of the system, an effort was made to give the user considerable freedom in data formatting and classification. This was stressed in the revised documentation.

A. A new program was prepared to provide summary tables of the output from the system. This was done primarily to enable output from the system to be used directly in the next End Results reports, thus, reducing the clerical effort required in preparing the report. However, due to the flexibility of the program, it can create abstracts of the System's

output for general use as well.

To improve program modification and storage, use of the DCRT Source Program Storage and Update System was incorporated into the maintenance process. At the suggestion of the End Results Section minor system changes were made to make the user control statements easier to understand.

Major Findings:

The initiation of an increased effort at system documentation encouraged the primary user, in this case the End Results Section, to view the output from this system which had been in constant use in recent years as a means to an end rather than an end in itself. Thus, the survival statistics generated by the system themselves became input to a summary report program, thereby making even greater use of the computer in analyzing large quantities of data.

Significance:

The documentation improvements of the system made its capabilities for handling a wide variety of survival analysis more apparent. Consequently, when the Surgery Branch of NCI began a survival analysis project involving data of a greatly different format and emphasis than that in use by the End Results group, the usefulness of the existing system could be demonstrated and was available for immediate use without any additional programming.

Proposed Course:

Continuing improvements to be made include an additional computation of medium survival time to be computed for each data classification within a given run and the addition of an option to plot the survival probabilities generated by the system.

Honors and Awards: None

Publications:

A number of monographs, reports, and articles have been published by the Cancer End Results Section, NCI and Surgery Branch, NCI in which results obtained from the Survival System are used.

Serial No. 7.30
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title:	NIH International Activities
Previous Serial Number:	None
Project Leader:	Judy Mahaffey, DCRT
Other Professionals:	None
Cooperating Units:	Fogarty International Center

Man Years:

Total:	0.20
Professional:	0.20
Others:	0.00

Project Description:

Objectives:

The purpose of this project is to design a system for the storage, maintenance and retrieval of information on NIH International Activities. All transactions on foreign research and training grants and contracts are now being maintained manually. The system, when complete, will be capable of maintenance and operation by the Fogarty Center.

Methods Employed:

The system is still in the design phase. Because many of the data elements needed to establish the International Activities data base currently exist in the DRG computer files, the design of the system will incorporate a method of extracting that data. Once the data base is established, a maintenance and reporting system will be developed to meet the Fogarty Center's 48-hour demand requirements.

Major Findings:

None

Significance:

None

Proposed Course:

When the design phase is complete and the system is operating, the complete system will be handed over to the Fogarty Institute.

Honors and Awards:

None

Publications:

None

Serial No. 7.31
1. Data Management Branch
2. Applied Systems Programming Section
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Medical Records: Diagnosis and Discharge

Previous Serial Number: None

Project Leader: Penny Brogan, DCRT

Other Professionals: None

Cooperating Units: Medical Records Department, CC

Man Years:

Total	0.6
Professional:	0.5
Others:	0.1

Project Description:

Objectives:

The purpose of this project is to design and implement a system for maintaining a computer file with basic diagnostic information on NIH patients. The file is composed of identifying information on each patient and the diagnosis rendered each time the patient is discharged. Various selective listings and statistics will be produced from the data.

Methods Employed:

Data will be collected, corrected and stored on WYLBUR. COBOL and IRS programs will be written to update the file and produce desired reports.

Significance:

The system will help the medical records librarians bring all their diagnostic data up-to-date. When the system is computerized it will be easier for the librarians to maintain the accuracy of the diagnostic records and to obtain composite lists as needed.

Proposed Course:

The system has been defined and is now being implemented.

Honors and Awards:

None

Publications:

None

Serial No. 7.32
1. Data Management Branch
2. Mathematical/Statistical Unit
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title:	Math/Stat Library
Previous Serial Number:	None
Project Leader:	George Shakarji, DCRT
Other Professionals:	Robert Baird, DCRT Kirk Dorn, DCRT Doug Gabriel, DCRT
Cooperating Units:	None

Man Years:

Total: 1.2
Professional: 1.2
Others:

Project Description:

Work in FY 1971:

1. A great deal of time is being spent writing detailed documentation and run instructions on all the mathematical/statistical programs that are fully tested and widely used. These documentations will comply with documentation standards set by the Data Management Branch, DCRT. The aim of this effort is to enable users to access programs without having to engage a professional programmer.

2. The Biomedical Computer Programs (BMD) from U.C.L.A. are undergoing vigorous testings under the newly obtained release 20 compilers. Since the BMD package contains its own I/O interface, different from that supplied by IBM, the testing of all the BMD programs, and the more recent BMD X-series programs is essential using both the source and object modules.

3. All IBM SSP source decks (main programs and subroutines) are now stored on 9-track 1600 bpi tape. A routine was developed that produces a listing and/or a punched card deck of any SSP routine residing on the tape. In the past the demand for source decks has been high. It is hoped that this procedure would save valuable programmer time.

4. To meet user demand, new programs were developed and added to the existing library programs.

   A. Analysis of Variance with repeated measures (the term repeated measures refers to a vector of observations for a given individual). For disproportionate group sizes there is an harmonic (unweighted means analysis) mean analysis option. A standardization option provides for the transformation of variables in different metrics to a vector of unitless observations.

   B. Factor Analysis using the centroid method.

      Factor Analysis achieved by successive reduction of the rank of the theoretical factor matrix by extraction of the centroid of the given correlation matrix. This extraction is based on a geometric interpretation of the factor matrix.

   C. Cross tabulations with tests of significance: the significance of "S" (in Kendall's tau and Goodman and Kruskal's gamma) and the information statistic.

   D. Interclass Correlation Coefficient-Analysis of variance.

      This Intraclass Correlation (ICC) program allows one to compute the ICC following a one-way or a two-way analysis of variance option. The assumptions underlying the validity of the ICC are those underlying the analysis of variance via normality, independence and homogeneity within variance.

      Two ICC coefficients appear in the output: (1) Reliability of Single Rating: This coefficient approximates the mean of the actual intercorrelations of the ratings (data, etc.) per person (Item, family, etc.). (2) Reliability of Average Rating (sometimes known as the Spearman-Brown prediction formula).

Significance:

The implementation and development of programs in the Math/Stat Library would give NIH investigators flexible,statistical and mathematical programs to be applied in the vast area of Biomedical Research.

Proposed Course:

To continue to develop and maintain a large program library and develop new procedures for storing library programs on tape or disk that would aid investigators in having easy access to and execution of library programs.

Honors and Awards: None

Publications: None

Serial No. 7.33
1. Data Management Branch
2. Mathematical/Statistical Unit
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Family Psychophysiology Project

Previous Serial Number: None

Project Leader: Dave Van Sant, DCRT

Other Investigators: None

Cooperating Units: Section on Personality Development, Adult Psychiatry Branch, NIMH

Man Years:

Total: 0.3
Professional: 0.3
Others: 0.0

Project Description:

Background:

The family psychophysiology project represents an extension of ongoing studies of family interaction and adolescent emotional disturbance conducted by the staff of the Section on Personality Development, NIMH.

Ordinarily, the presence of anxiety in family interaction is determined from patients' self-reports and the therapists' inferences drawn from observable family behavior. To these indicators a psychophysiologic variable, the galvanic skin response has now been added.

Objectives:

The objective of this project, as stated by the Section on Personality Development, is to study the relationship between subjective experience and behavior within the family, and permit the examination of the process of therapeutic intervention in families of troubled adolescents.

Work in F.Y. 1971:

Thus far, the programming effort of this project has accomplished

the following tasks:

1) Edit and combine data, matching information from two sources, matching each individual skin response with the speech unit in which it occurred. The conductance value of each response amplitude is computed, as well as an amplitude per second value.

2) Descriptive statistical computations were performed on such factors as the rate of response as well as total amplitude, number of responses and length of speech unit.

Major Findings:

At this stage of the project only limited evaluations have been made. Therefore, major findings are not available at this time.

Proposed Course:

Continue evaluations of the data.

Honors and Awards: None

Publications: None

Serial No. 7.34
1. Data Management Branch
2. Mathematical/Statistical Unit
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Cognition and Identity Project

Previous Serial Number: None

Project Leaders: Ray Danner, DCRT
Dave Van Sant, DCRT

Other Investigators: None

Cooperating Units: Adult Psychiatry Branch, NIMH

Man Years:

Total: 0.2
Professional: 0.2
Others:

Project Description:

Background:

This is a longitudinal study of individual development in several contrasting adolescent samples: lower-socioeconomic class white and black boys; middle-class patients and non-patients. The control concept to the project is that of "identity formation", a notion widely used to describe a major adolescent developmental step.

Objectives:

Develop programs for the following tasks:

1. To derive more precise measurement of identity developments, and associated cognitive processes.

2. To study and compare these important processes in diverse and contrasting populations.

3. To explore relationships among identity formation patterns and cognitive patterns within individuals and between specific groups.

Work in F.Y. 1971:

Programs were developed to compute:

1. The basic analysis of self-image inter-correlations at single points and between testings.

2. The information statistics, H and T.

3. Inter-correlations of the segments over time, and tests for their significance.

4. Analysis of the segment averages, and conversion of these into a standard "D index".

5. Transformations of skewed data to normal distributions and producing histograms of each set of the data.

6. Analysis of variance for the information statistics.

Major Findings:

No major findings are available at this time since this study is not complete.

Proposed Course:

To develop computer programs that would further define indices in identity formation.

Honors and Awards: None

Publications: None

Serial No. 7.35
1. Data Management Branch
2. Mathematical/Statistical Unit
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Professional Values and Nurse-Doctor-Patient Interaction

Previous Serial Number: None

Project Leaders: Dave Van Sant, DCRT
Doug Gabriel

Other Investigators: None

Cooperating Units: Experimental Group and Family Studies Section, Adult Psychiatry Branch, NIMH

Man Years:

Total: 0.25
Professional: 0.25
Others: 0.00

Project Description:

Background:

This project represents an attempt to study the interaction between nurses, doctors and patients on a psychiatric ward. Data consisted of scores on questionnaires and observations which were obtained experimentally by forming triads among nurse, doctor and patient. Each triad was given an experimental task and verbal and non-verbal behavior were assessed.

Objectives:

The Experimental Group and Family Studies Section, NIMH, hypothesized in this study that nurses and doctors, who value social openness in relationships and value treatments that depend substantially on a variety of group therapies, would interact directly and openly with each other and be able to effectively integrate a patient into their group.

Work in F.Y. 1971

All the scores were edited and erroneous data were corrected. The

data analyzed consisted of analysis of variance comparing triads. In addition responses to the questionnaires were factor analyzed and a variety of correlational studies were done comparing questionnaire responses and performance in the experimental situations.

Major Findings:

The statistical analysis, showed not only that professional values have a substantial effect on the style of nurse-doctor-patient interaction but also that this effect is, in many cases, modified by the nurse's personality structure.

Significance:

Significance for mental health: This study provides subjective data indicating kinds of professional training and personnel selection required to produce particular kinds of psychiatric treatment and programs on in-patient services.

Honors and Awards: None

Publications: None

Serial No. 7.36
1. Data Management Branch
2. Mathematical/Statistical Unit
3. Bethesda

PHS-NIH
Individual Project Report
July 1, 1970 through June 30, 1971

Project Title: Internatioanl Pilot Study of Schizophrenia

Previous Serial Number: None

Project Leaders: Fred Yamada, DCRT
Kirk Dorn, DCRT

Other Investigators: None

Cooperating Units: Adult Psychiatric Branch, Psychiatric Assessment Section, NIMH

Man Years:

Total:	0.8
Professional:	0.8
Others:	0.0

Project Description:

Background:

DCRT has contributed to three of the major projects being carried on by the Psychiatric Assessment Section. The Psychiatric Assessment Section, as part of the Adult Psychiatry Branch, NIMH, is a collaborating center in the World Health Organization sponsored International Pilot Study of Schizophrenia. This study involves the research efforts of psychiatric centers in nine different countries in co-ordination with WHO Headquarters in Geneva to develop methods for evaluating psychiatric patients. Such methods that can be applied in different countries and cultural settings will form the basis of future studies of psychiatric epidemiology. A major part of this study is the use of standardized interview forms in evaluating more than 1200 patients in the nine different countries. The level of agreement among raters has had to be measured to evaluate whether raters from different backgrounds could use the interview forms in similar ways. The Psychiatric Assessment Section has been asked to evaluate the reliability of ratings between raters and between centers.

Work for F.Y. 1971:

1. Programs were developed to analyze and solve the complex task of

measuring the different rates of reliability among raters from the same centers, among raters with different levels of knowledge of English, among raters from different centers, and among raters trained in different schools of psychiatry. The results of this work will appear in Volume one of the IPSS report.

2. To develop and evaluate the rating schedules for the International Pilot Study of Schizophrenia, it has been essential to evaluate whether rating bias existed on a significant level either for individual raters or particular psychiatric centers. The nature of the data and the magnitude of the study required that new measures for evaluating bias be developed. In conjunction with DCRT it was possible to develop measures of bias and use them to analyze the data of the IPSS. The results of this work will appear in Volume one of the IPSS report.

3. A major goal of the Psychiatric Assessment Section has been to develop new criteria for diagnosis of psychiatric patients that will produce more valid diagnostic groups than exist currently. One method of special interest has been the use of clustering techniques applied to psychiatric taxonomy. Since clustering methodology is still at a primitive level, it has been necessary to evaluate several different clustering programs in analyzing data obtained from psychiatric patients seen by the staff of the Psychiatric Assessment Section. Modification of existing clustering techniques was also necessary. These efforts were carried out with considerable assistance from DCRT. Difficult and creative programming was required to apply clustering techniques to psychiatric data in a way that produced useful results. Considerable progress has been made in this effort and two papers based on the application of the techniques to psychiatric data have been submitted for publication.

Methods Employed:

Statistical methods have been developed to measure different rates of reliability among raters from the same centers, among raters with different levels of knowledge of English, among raters from different centers, and among raters trained in different schools of psychiatry. Also extensive clustering techniques were applied to psychiatric taxonomy in addition to standard statistical techniques.

Major Findings:

In conjunction with DCRT it was possible to develop measures of bias and use them to analyze the data of the IPSS. The analysis is continuing and findings will appear in IPSS reports.

Significance:

These studies will help develop methods for evaluating psychiatric patients in different psychiatric centers in different countries.

Proposed Course:

Continue evaluations of the data.

Honors and Awards: None

Publications: None